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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.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

Associate editors
Vladimir A Fomichev, K. E. Tsitsolkovsky Russian State Tech Univ, Russia; Olga S Fomicheva, Studio "Culture, Ecology, and Foreign Languages", Russia; Piet Kammer, 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.

Advisory board
Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J. R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Ruddy Lelouche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGeer, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Riechiro 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; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan.

Assistant Editors
Yuan-Ihsuan (Karen) Lee, National Chiao Tung University, Taiwan.

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

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

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

Abstracting and Indexing

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

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

For detailed information on how to format your submissions, please see:
http://www.ifets.info/guide.php

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

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

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

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

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

## Special issue articles

- **Guest Editorial – Technology Supported Cognition and Exploratory Learning**  
  *Dirk Ifenthaler, Pedro Isaias, Kinshuk, Demetrios G. Sampson and J. Michael Spector*  
  1-1

- **Epistemological Beliefs and Ill-structured Problem-solving in Solo and Paired Contexts**  
  *Charoula Angeli and Nicos Valanides*  
  2-14

- **A Study on Exploiting Commercial Digital Games into School Context**  
  *Hercules Panoutsopoulos and Demetrios G. Sampson*  
  15-27

- **Aberrance Detection Powers of the BW and Person-Fit Indices**  
  *Tsai-Wei Huang*  
  28-37

- **Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios**  
  *Dirk Ifenthaler*  
  38-52

- **Presence and Middle School Students’ Participation in a Virtual Game Environment to Assess Science Inquiry**  
  *Catherine C. Schifter, Diane Jass Ketelhut and Brian C. Nelson*  
  53-63

## Full length articles

- **Any Effects of Different Levels of Online User Identity Revelation?**  
  *Fu-Yun Yu*  
  64-77

- **Relationship between Students’ Emotional Intelligence, Social Bond, and Interactions in Online Learning**  
  *Heeyoung Han and Scott D. Johnson*  
  78-89

- **A Folksonomy-based Guidance Mechanism for Context-aware Ubiquitous Learning: A Case Study of Chinese Scenic Poetry Appreciation Activities**  
  *Wen-Chung Shih, Shian-Shyong Tseng, Che-Ching Yang, Chih-Yu Lin and Tyne Liang*  
  90-101

- **How Concept-mapping Perception Navigates Student Knowledge Transfer Performance**  
  *Kuo-Hung Tseng, Chi-Cheng Chang, Shi-Jer Lou, Yue Tan and Chien-Jung Chiu*  
  102-115

- **Elementary School Students’ Perceptions of the New Science and Technology Curriculum by Gender**  
  *Mehmet Nuri Gömleksiz*  
  116-126

- **Student Satisfaction, Performance, and Knowledge Construction in Online Collaborative Learning**  
  *Chang Zhu*  
  127-136

- **Design of a Motivational Scaffold for the Malaysian e-Learning Environment**  
  *Nor Aziah Alias*  
  137-151

- **Understanding of the Relationship Between Interest and Expectancy for Success in Engineering Design Activity in Grades 9-12**  
  *Oenardi Lawanto, Harry B. Santoso and Yang Liu*  
  152-161

- **An Investigation into Parent-Child Collaboration in Learning Computer Programming**  
  *Janet Mei-Chuen Lin and Shu-Fen Liu*  
  162-173

- **Examination of Co-construction of Knowledge in Videotaped Simulated Instruction**  
  *Sitikiye Kuter, Zehra Altinay Gazi and Fahriye Altinay Aksal*  
  174-184

- **Prospective EFL Teachers’ Perceptions of ICT Integration: A Study of Distance Higher Education in Turkey**  
  *Murat Hismanoglu*  
  185-196

- **The Impact of Recurrent On-line Synchronous Scientific Argumentation on Students’ Argumentation and Conceptual Change**  
  *Chien-Hsien Chen and Hsiao-Ching She*  
  197-210
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyzing the Learning Process of an Online Role-Playing Discussion Activity</td>
<td>211-222</td>
</tr>
<tr>
<td>Huei-Tse Hou</td>
<td></td>
</tr>
<tr>
<td>A Context-Aware Mobile Learning System for Supporting Cognitive Apprenticeships in Nursing Skills Training</td>
<td>223-236</td>
</tr>
<tr>
<td>Po-Han Wu, Gwo-Jen Hwang, Liang-Hao Su and Yueh-Min Huang</td>
<td></td>
</tr>
<tr>
<td>Rebeka Lukman and Majda Krajnc</td>
<td></td>
</tr>
<tr>
<td>Learning Achievement in Solving Word-Based Mathematical Questions through a Computer-Assisted Learning System</td>
<td>248-259</td>
</tr>
<tr>
<td>Tzu-Hua Huang, Yuan-Chen Liu and Hsiu-Chen Chang</td>
<td></td>
</tr>
<tr>
<td>Patterns of Interaction and Participation in a Large Online Course: Strategies for Fostering Sustainable Discussion</td>
<td>260-272</td>
</tr>
<tr>
<td>Jiyeon Lee</td>
<td></td>
</tr>
<tr>
<td>A Fuzzy Logic-based Personalized Learning System for Supporting Adaptive English Learning</td>
<td>273-288</td>
</tr>
<tr>
<td>Tung-Cheng Hsieh, Tzone-I Wang, Chien-Yuan Su and Ming-Che Lee</td>
<td></td>
</tr>
<tr>
<td>Shared Mental Models on the Performance of e-Learning Content Development Teams</td>
<td>289-297</td>
</tr>
<tr>
<td>Il-Hyun Jo</td>
<td></td>
</tr>
<tr>
<td>Intelligent Discovery for Learning Objects Using Semantic Web Technologies</td>
<td>298-312</td>
</tr>
<tr>
<td>I-Ching Hsu</td>
<td></td>
</tr>
<tr>
<td>A Model for Predicting Learning Flow and Achievement in Corporate e-Learning</td>
<td>313-325</td>
</tr>
<tr>
<td>Young Ju Joo, Kyu Yon Lim and Su Mi Kim</td>
<td></td>
</tr>
<tr>
<td>Providing Adaptivity in Moodle LMS Courses</td>
<td>326-338</td>
</tr>
<tr>
<td>Marijana Despotović-Zrakić, Aleksandar Marković, Zorica Bogdanović, Dušan Barać and Srdjan Krčo</td>
<td></td>
</tr>
<tr>
<td>Agent Prompts: Scaffolding for Productive Reflection in an Intelligent Learning Environment</td>
<td>339-353</td>
</tr>
<tr>
<td>Longkai Wu and Chee-Kit Looi</td>
<td></td>
</tr>
<tr>
<td>Utilizing a Collaborative Cross Number Puzzle Game to Develop the Computing Ability of Addition and Subtraction</td>
<td>354-366</td>
</tr>
<tr>
<td>Yen-Hua Chen, Chee-Kit Looi, Chiu-Pin Lin, Yin-Juan Shao and Tak-Wai Chan</td>
<td></td>
</tr>
<tr>
<td>Effects of Speech-to-Text Recognition Application on Learning Performance in Synchronous Cyber Classrooms</td>
<td>367-380</td>
</tr>
<tr>
<td>Wu-Yuin Hwang, Rustam Shadiev, Tony C. T. Kuo and Nian-Shing Chen</td>
<td></td>
</tr>
<tr>
<td>Teachers' Belief and Use of Interactive Whiteboards for Teaching and Learning</td>
<td>381-394</td>
</tr>
<tr>
<td>Yalin Kilic Turel and Tristan E. Johnson</td>
<td></td>
</tr>
</tbody>
</table>
Guest Editorial - Technology Supported Cognition and Exploratory Learning

Dirk Ifenthaler¹, Pedro Isaias², Kinshuk³, Demetrios G. Sampson⁴ and J. Michael Spector⁵

¹University of Mannheim, Germany // ²Portuguese Open University, Portugal // ³Athabasca University, Canada // ⁴University of Piraeus & CERTH, Greece // ⁵University of North Texas, USA // ifenthaler@uni-mannheim.de // pisaias@uab.pt // kinshuk@athabascau.ca // sampson@iti.gr // Mike.Spector@unt.edu

The International Association for the Development of the Information Society (IADIS; see http://www.iadis.org/) 2010 International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA) was hosted by the "Politehnica" University of Timisoara, Romania in October 2010 (see http://www.iadis.org/ceda2010/). The IADIS CELDA 2010 conference aims to address the main issues concerned with evolving learning processes and supporting pedagogies and applications in the digital age. There have been advances in both cognitive psychology and computing that have affected the educational arena. The convergence of these two disciplines is increasing at a fast pace and affecting academia and professional practice in many ways. Paradigms such as just-in-time learning, constructivism, student-centered learning and collaborative approaches have emerged and are being supported by technological advancements such as simulations, virtual reality and multi-agents systems. These developments have created both opportunities and areas of serious concerns.

Editors of this special issue selected a number of papers presented at IADIS CELDA 2010 conference that were very highly rated by reviewers, well received at the conference, and nicely complementary in terms of research, theory, and implications for learning and instruction. These papers have been edited and revised based on feedback from conference participants and subsequent review by the editors of this special issue and reviewers recruited to assist in this process. The organizing committee of IADIS CELDA 2010 proposed a special issue of Educational Technology & Society Journal based on selected papers from IADIS CELDA 2010. The result is the five papers included in this special issue.

The first paper in this special issue is “Epistemological Beliefs and Ill-Structured Problem-Solving in Solo and Paired Contexts”, authored by Charoula Angeli (University of Cyprus, Cyprus) and Nicos Valanides (University of Cyprus, Cyprus), examines the relationship between epistemological beliefs and quality of thinking when participants first thought about an ill-structured problem alone, and then with another person in a dyad.

In the second paper, “A Study on Exploiting Commercial Digital Games into School Context”, authored by Hercules Panoutsopoulos (University of Piraeus & Doukas School, Greece) and Demetrios G. Sampson (University of Piraeus & CERTH, Greece), examines the effect of a general-purpose commercial digital game (namely, the “Sims 2-Open for Business”) on the achievement of standard curriculum Mathematics educational objectives as well as general educational objectives as defined by standard taxonomies.

In the third paper, “Aberrance Detection Powers of the BW and Person-Fit Indices”, authored by Tsai-Wei Huang (National Chiayi University, Taiwan), presents a study that compared the aberrance detection powers of the BW person-fit indices with other group-based indices (SCI, MCI, NCI, and Wc&Bs) and item response theory based (IRT-based) indices (OUTFITz, INFITz, ECI2z, ECI4z, and Iz).

In the fourth paper, “Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios”, authored by Dirk Ifenthaler (University of Mannheim, Germany), reports an experimental study with 98 participants where effective instructional interventions for self-regulated learning within problem-solving processes are investigated.

In the fifth paper, “Presence and Middle School Students’ Participation in a Virtual Game Environment to Assess Science Inquiry”, authored by Catherine C. Schifter (Temple University, USA), Diane Jass Ketelhut (Temple University, USA) and Brian C. Nelson (Arizona State University, USA), introduces a project to design and implement a virtual environment (SAVE Science) intended to assess (not teach) middle school students’ knowledge and use of scientific inquiry through two modules developed around curriculum taught in middle schools in Pennsylvania, USA.
Epistemological Beliefs and Ill-structured Problem-solving in Solo and Paired Contexts

Charoula Angeli and Nicos Valanides
Department of Education, University of Cyprus, 11-13 Dramas street, Nicosia 1678, Cyprus // cangeli@ucy.ac.cy // nichri@ucy.ac.cy

ABSTRACT
A mixed-method exploratory approach was employed to examine the relationship between epistemological beliefs and quality of thinking when participants first thought about an ill-structured problem alone, and then with another person in a dyad. The results showed that there was not a systematic connection between epistemological beliefs and ill-structured problem solving in either solo or paired contexts. It is speculated that the emotional and cultural nature of the problem affected participants’ problem-solving approach. It is recommended that future empirical studies examine the relationship between epistemological beliefs and thinking in a contextualized way by assuming an integrative approach so that emotions, epistemological beliefs, and cognition are considered systemically.

Keywords
Epistemological beliefs, Individual thinking, Paired thinking, Ill-structured problem solving.

Introduction
A lot of the problems that we are often confronted with, either in our personal lives or in the workplace, are mostly ill-structured, that is, problems for which there is real uncertainty as to how they can best be solved. According to Jonassen (1997), ill-structured problems are unique interpersonal activities and require learners to express personal beliefs; thus, for this reason, cognitive processes alone are insufficient requirements for solving ill-structured problems, because epistemological beliefs affect the ways that learners naturally tend to approach these problems (Oh & Jonassen, 2007; Mandler, 1989; Rogoff, 1990, 2003). We use the term epistemological beliefs to refer to beliefs about the nature of knowledge (certainty of knowledge) and knowing (source of knowledge and justification of knowledge) (Hofer, 2001).

Empirical findings showed that epistemological beliefs affect reasoning about ill-structured problems (Bendixen & Schraw 2001; Schommer & Dunnell, 1997; Sinatra, Southerland, McConaughy, & Demastes, 2003; Schraw, 2001). Research in this area, however, has not addressed closely the role of social context on one’s epistemological beliefs. In other words, could it be possible for Bendixen, Schraw, Schommer, and Dunnell to obtain different results about the role of epistemological beliefs on students’ reasoning had they asked their students to think about an ill-structured problem, not alone, but with others in a collaborative setting? Do epistemological beliefs behave the same way when one thinks about a problem individually or with others in a group?

Therefore, to remedy for the lack of research on the role of context on epistemological beliefs, in this study we considered socio-cultural aspects of the problem-solving context, and assumed a mixed-method exploratory approach in order to better understand how learners with naïve or sophisticated epistemological beliefs think about an ill-structured controversial problem individually or in dyads.

Literature review
Jonassen (1997) distinguished well-structured from ill-structured problems, and articulated differences in cognitive processing engaged by each. Ill-structured problem solving often requires solvers to consider multiple perspectives and apply several criteria while evaluating problems or solutions. The ability to do so depends partially on solvers underlying beliefs about knowledge and how it develops. Since ill-structured problems have commonly divergent or alternative solutions, solvers must develop justification or an argument for supporting the rationale of their selection of a particular solution (Voss & Post, 1988).
For ill-structured problems, the process of justification requires the identification of as many as possible of the various perspectives, supporting arguments and evidence on opposing perspectives, evaluating information, and developing and arguing for the best possible solution (Voss & Means, 1991). According to Churchman (1971), reconciling different interpretations of phenomena based on solvers’ goals or perceptions about the nature of the problem is a critical process in developing justification. Thus, the solver’s epistemic cognition is an important component in order to develop justification for ill-structured problems (Kitchener, 1983). For developing justification, individuals need epistemic cognition in order to understand that ill-structured problems do not always have a correct solution, and how to choose between alternative solutions. The process of developing justification though for well-structured problems is quite different and focuses mostly on the development of a logical argument in support of the correct solution. Overall, research findings have consistently shown that performance in solving well-structured problems is independent of performance on ill-structured tasks, with ill-structured problems engaging a different set of epistemological beliefs, and thus a different process for developing justification about the problem at hand (Schraw, Dunkle, & Bendixen, 1995; Hong, Jonassen, & McGee, 2003; Jonassen & Kwon, 2001).

Research concerning the role of epistemological beliefs in learning can be traced back to the work of Perry (1970). In Perry’s work, undergraduates during their freshman year in college tended to believe in simple, certain knowledge that is handed down by authority, but by the time they reached their senior year in college they changed their beliefs and believed more in tentative, complex knowledge derived from reason. In a comprehensive review of the literature, Hofer and Pintrich (1997) and Hofer (2001) stated that researchers whose work is based on Perry’s, such as for example Baxter Magolda (1992), Belenky, Clinchy, Goldberger, and Tarule (1986), and King and Kitchener (1994), studied personal epistemology in terms of stage models. Other researchers, such as Schommer (1990, 1998) suggested a different conceptualization of epistemological development and proposed that personal epistemology would be better conceptualized as a system of independent beliefs that do not necessarily develop at the same rate and time. Essentially, Schommer’s (1990, 1998) work distinguished between unidimensional and multidimensional models of epistemological development proposing that probably not all beliefs develop at the same rate. This approach suggests that, at some point in time, an individual may come to believe that knowledge is highly interrelated, but yet also believe that knowledge is certain. In particular, Schommer (1994) proposed a taxonomy of five dimensions of epistemological beliefs, namely, (a) beliefs about the stability of knowledge, ranging from tentative to unchanging, (b) beliefs about the structure of knowledge, ranging from isolated bits to integrated concepts, (c) beliefs about the source of knowledge, ranging from handed down by authority to assemble from observation and reason, (d) beliefs about the speed of knowledge acquisition, ranging from quick-all-or-none learning to gradual learning, and (e) beliefs about the control of knowledge acquisition, ranging from fixed at birth to lifelong improvement. In order to assess these beliefs, Schommer (1990) constructed a questionnaire with 63 Likert-type items on a scale from 1 (strongly disagree) to 5 (strongly agree).

A different approach to studying personal epistemology was proposed by King and Kitchener (1994). King and Kitchener took into consideration the contextual dependencies of students’ beliefs about knowledge and constructed a seven-stage rubric to assess two aspects of epistemological beliefs, namely, “view of knowledge” and “justification of beliefs.” In this scheme, a score ranges from 1 to 7 and indicates the epistemological stage of the individual. The first three sub-stages of the rubric constitute the pre-reflective stage. During these sub-stages one’s view of knowledge progresses from absolute certainty to absolute certainty that can be temporarily uncertain and not immediately available, to absolute certainty about some things and temporary uncertainty about other things. The concept of justification during these sub-stages changes from no need for justification to justification via authorities or direct observation, to justification via authorities in some areas and via what one feels right at instances where knowledge is uncertain. The fourth and fifth sub-stages constitute the quasi-reflective stage. During these sub-stages one’s view of knowledge progresses from no certainty because of situational variables, such as time, to no certainty except via personal perspectives within a specific context. The concept of justification during these sub-stages shifts from justification via idiosyncratic evaluations of evidence and unevaluated beliefs to justification via rules of inquiry for a particular context. In essence, quasi-reflective thinking recognizes that one cannot always know with certainty and that thinking is contextual and relative to one’s experiences. The sixth and seventh sub-stages constitute the reflective judgment stage. During these sub-stages, one’s view of knowledge progresses from some personal certainty about beliefs, based on evaluations of evidence on different sides of the question, to certainty that some knowledge claims are better or more complete than others although they always remain open to reevaluation. The concept of justification emerges from justification via generalized rules of inquiry, personal evaluations that apply across contexts and evaluated views of experts, to justification via reasonable conjectures about reality or the world.
based on an integration and evaluation of data and/or opinions. Therefore, during the reflective judgment stage, one is able to engage in rational inquiry and derive a reasoned judgment.

Research findings established a relationship between epistemological beliefs and reasoning about ill-structured problems (Bendixen & Schraw 2001; Bendixen, Dunkle, & Schraw, 1994; Bendixen, Schraw, & Dunkle, 1998; Schommer & Dunnell, 1997; Sinatra, Southerland, McConaughy, & Demastes, 2003). For example, research by Bendixen, Dunkle, and Schraw (1994) showed that students who view ability as innate and thus fixed may be less inclined to develop and use advanced reasoning skills when thinking about ill-structured issues. Also research by Schraw, Dunkle, and Bendixen (1995) found that well-structured and ill-structured problems engaged different epistemological beliefs. Schommer and Dunnell (1997) found that the more students believed that the ability to learn is fixed at birth, that learning is quick or not-at-all, and that knowledge is unchanging, the more likely they were to write overly simplistic solutions to problems.

A question that could be asked at this point is whether different research results could be obtained if socio-cultural aspects of learning were taken into consideration in the aforementioned research studies. We believe this is a very important question that needs to be investigated, because more and more people form partnerships and think with others, and thus it has become widely accepted that cognitions are not decontextualized tools and products of mind, but situated and distributed. In spite of this, research has not addressed closely the role of social context on one’s epistemological beliefs. Therefore, in this study, we aimed to remedy for this lack of research in the current body of literature and sought to answer the following questions:

1. What are the elements of students’ reasoning when thinking about an ill-structured problem in solo and paired problem-solving contexts?
2. What is the relationship between students’ epistemological beliefs and students’ reasoning when thinking about an ill-structured problem in solo and paired problem-solving contexts?

Methodology

Participants

Twenty graduate students from a teacher education department volunteered to participate in this small scale exploratory study. The majority of the participants were females. The average age of the participants was 23.5 years.

Instruments

Epistemological beliefs questionnaire

The Epistemological Beliefs Questionnaire (EBQ) was used for collecting data. The EBQ consisted of five questions measuring participants’ perceived importance of an ill-structured issue (question 1) and their epistemological beliefs (questions 2-5). The questions on the EBQ were adapted from King and Kitchener’s (1994) interview questions. Question 1 was a Likert-type question with ratings from 7 (extremely important) to 1 (not important at all). The question was simply asking each student to rate the importance of a complex geopolitical issue regarding the reunification of Cyprus on the basis of the Annan Plan. Question 2 prompted the participants to state how it was possible for two experts to express different points of view on the issue, and explain how this sort of disagreement among experts could happen. Analogously, question 5 raised the issue of whether it could be known for sure that an individual's position on a specific issue at hand was correct, and participants were asked to explain their answer accordingly. Questions 3 and 4 dealt with how beliefs should be justified. Specifically, question 3 prompted the participants to explain how two experts could justify different views on the same issue, and question 4 asked the participants to explain how they themselves came to form and justify their point of view on the issue.

The data collected with the EBQ were analyzed with a three-stage rubric, shown in Table 1, which constituted an adaptation of King and Kitchener’s (1994) seven-stage model to a simpler version. Specifically, we reduced the original seven-stage model to a three-stage scheme because, “generally speaking, models of epistemological development postulate three broad stages characterized first by absolutist beliefs, followed by the advent of relativist beliefs, followed by the advent of pluralist beliefs in which beliefs are viewed as relative, but more or less defensible
depending on one’s ability to support them with warrants” (Schraw, 2001, p. 456). There is also unanimous agreement that regardless of the number of epistemological beliefs there are fundamentally two different types corresponding to “view of knowledge” and “justification of beliefs” (Hofer & Pintrich, 1997).

<table>
<thead>
<tr>
<th>ABSOLUTIST THINKING</th>
</tr>
</thead>
<tbody>
<tr>
<td>View of knowledge:</td>
</tr>
<tr>
<td>Knowledge is assumed to be either right or wrong. If it is not absolutely certain, it is only temporarily uncertain and will soon be determined. A person can know with certainty through three sources: (a) direct observation; (b) what “feels right;” and (c) authorities (experts, teachers, parents).</td>
</tr>
</tbody>
</table>

| Concept of justification: |
| Beliefs need no justification or they are justified through an authority figure such as a teacher or a parent. Most questions are assumed to have a right answer so there is little or no conflict in making decisions about disputed issues. |

<table>
<thead>
<tr>
<th>RELATIVIST THINKING</th>
</tr>
</thead>
<tbody>
<tr>
<td>View of knowledge:</td>
</tr>
<tr>
<td>Knowledge is uncertain (there is no right or wrong) and idiosyncratic to the individual. Knowledge is seen as subjective and contextual.</td>
</tr>
</tbody>
</table>

| Concept of justification: |
| Beliefs are justified by giving reasons and evidence idiosyncratic to the individual. Beliefs are filtered through a person’s experiences and criteria for judgment. |

<table>
<thead>
<tr>
<th>REFLECTIVE THINKING</th>
</tr>
</thead>
<tbody>
<tr>
<td>View of knowledge:</td>
</tr>
<tr>
<td>Knowledge is constructed by comparing evidence and opinion on different sides of an issue. Knowledge is the outcome of the process of reasonable inquiry leading to a well-informed understanding.</td>
</tr>
</tbody>
</table>

| Concept of justification: |
| Beliefs are justified by comparing evidence and opinion from different perspectives. Conclusions are defended as representing the most complete, plausible understanding of an issue on the basis of the available evidence. |

Thus, we collapsed the three pre-reflective sub-stages into one stage, that of absolutist thinking, the two quasi-reflective sub-stages into another stage, namely, relativist thinking, and the last two sub-stages into a third stage, that of reflective thinking.

In scoring the responses on the EBQ, two raters were trained using the same procedures that we established in a previous study about the development of epistemological beliefs (Valanides & Angeli, 2005). Succinctly, the two trained raters independently scored the answers to the four questions of the EBQ. Each of the three stages of the rubric consisted of two subsections: (1) view of knowledge, and (2) justification of beliefs. Responses to each one of the four questions were rated by referring to one of the subsections of each stage. For example, the responses to questions 2 and 5 were rated based on their fit with the section on view of knowledge of each stage, whereas the responses to questions 3 and 4 were rated based on their fit with the section on justification of beliefs of each stage.

Each question of the EBQ was analyzed and scored using a scale from 1 to 3. These numbers corresponded to the three stages of the simplified three-stage model of epistemological beliefs shown in Table 1. A score of 1, 2, and 3 indicated performance at the level of absolutist thinking, the level of relativist thinking, and the level of reflective thinking, respectively. Scores were then summarized into a four-digit code indicating the respective scores from each of the four questions. For example, 1121 indicated performance at the level of absolutist thinking for questions 2, 3, and 5, and at the relativist thinking for question 4. As King and Kitchener (1994) argue, this scoring procedure is
based on the assumption that no single-stage score best represents a participant’s response and, furthermore, it allows for subject variability seen in his or her responses to be reflected in the overall rating of the problem. Then, the ratings returned by each rater (first-round ratings) were compared, and if there were questionnaires on which raters’ mean scores differed by half a stage or more, then they were given back to the raters to blindly re-evaluate them (second-round ratings). If there were any further discrepancies after the second round, then the two raters discussed their disagreements and consensus ratings were finally assigned. A mean score for each student was finally derived by averaging the scores from both raters.

Research procedures

Data were collected in two research sessions. The first research session lasted 90 minutes. During the first 30 minutes, participants completed the EBQ. Then, they were given 20 minutes to read some materials about the history of Cyprus and the issue of the reunification of Cyprus on the basis of the Annan plan. Subsequently, for the next 40 minutes, students had to work individually in order to write their position on the reunification of Cyprus using a computer tool that was developed specifically for the purposes of this study.

Participants’ positions were saved in log files that were later downloaded for analysis purposes. Written instructions asked the participants to analyze the issue broadly from different perspectives, and support their position with reason and evidence. The reading materials were available to the participants throughout the session.

Seven days later, the second research session took place. We considered the seven days between the first and second research sessions to be enough elapsed time, even though we do acknowledge the fact that the order might have made some difference. During the second research session, which lasted 40 minutes, participants were randomly assigned into dyads. Students in each dyad were given the same instructions as in the first session with the only difference that they were instructed to discuss the issue regarding the reunification of Cyprus on the basis of the Annan Plan using a synchronous text-based computer-supported collaborative environment that was specifically designed and developed for the purposes of this study. The transcript from each dyad was saved into a log file that was downloaded later for analysis. Partners in each dyad were anonymous and were placed in two different rooms to eliminate physical contact between them. As in the first research session, participants could also use the reading materials any time they needed them.

Data analysis

Log files for individual thinking were downloaded and analyzed by two raters independently. Each rater used Inspiration, a computer software, to diagram the flow of reasoning as reported in each log file. A scheme with scoring rules was provided to the two raters. Specifically, the scoring scheme included criteria important to good quality thinking such as the extent to which there was a point of view that was clearly supported, explanations, opposing arguments, and discussion of an alternative point of view and reasons for supporting it. Each rater followed this inductive approach to create different categorizations of the quality of thinking and to classify each log file in one of these categories. Inter-rater reliability was computed using the percentage of agreement between the two raters in terms of classifying a log file in the same category and was found to be 88%. The raters and researchers discussed observed disagreements and easily resolved the existing differences.

Regarding paired thinking, the transcripts from the collaborative sessions were also downloaded and analyzed by the two raters using Inspiration. This analysis focused on the individual contributions to the dialogue and the exchanges between the two partners including the number of disagreements between them. Inter-rater reliability was computed and was found to be 90%.

In order to investigate the differences between solo and paired thinking, the transcripts were analyzed using a coding scheme that was developed through a grounded theory approach (Strauss & Corbin, 1990). The first version of the coding scheme was inductively constructed by the two researchers and it was then given to an independent rater for confirmation. The independent rater and the researchers then discussed all discrepancies and an improved version of the coding scheme was prepared. The two other independent raters analyzed all solo and paired transcripts and a
Pearson $r$ between the two ratings was calculated and found to be 0.83, which was regarded satisfactory considering the complexity of the data.

**Results**

According to EBQ scores, nine participants were found to be reflective thinkers (i.e., stage 3 of epistemological development), and 11 of them relativist thinkers (i.e., stage 2 of epistemological development). None of the participants was found to be at stage 1 of epistemological development. Those who scored at stage 3 were classified as “High Epistemological Beliefs” and those who scored at stage 2 were classified as “Low Epistemological Beliefs”. Participants were then randomly assigned to ten dyads, seven of which were High/Low dyads, two were Low/Low dyads, and one was High/High.

**Individual thinking**

Participants’ individual transcripts were downloaded and analyzed using a diagrammatic technique. The diagrammatic technique was used to visualize the flow of participants’ reasoning as it appeared in the transcripts. Four types of diagrams, namely, Type A, B, C, and D shown in Figures 1, 2, 3, and 4 emerged from this analysis.

Diagram type A thinking (shown in Fig. 1) shows low-level thinking depicting failure to think about the problem systematically. Instead, several points of view are expressed in a disconnected way without any consistent flow of logic. Of the 20 participants, three of them fell into this category, namely, P116, P96, and P111. P116 scored high on the EBQ and the others low.

![Diagram A](image)

*Figure 1. Diagram type A thinking*

Diagram type B thinking (shown in Fig. 2) shows thinking that is reasoned within a stated point of view supported by a number of reasons. The flow of logic is well-organized and systematic. There is breadth in thinking but not depth, as the arguments presented are not elaborated adequately. Also, the thinking appears to be monological. Succinctly, monological thinking is thinking that hardly ever considers major alternative points of view, or hardly ever responds to objections framed by opposing views (Paul, 1995). The majority of the participants fell into this category, i.e., P107, P103, P113, P104, P114, P108, P115, P109, P110, P95, P106, and P112. Of these participants, six scored high on the EBQ and six low.

Diagram type C thinking (shown in Fig. 3) shows depth and breadth in thinking. However, as it was the case with diagram type B thinking, the thinking appears to be monological as different points of view or opposing arguments are not examined. Three participants exhibited thinking in this category (P100, P102, and P99). P102 scored high on the EBQ, and the others low.
Diagram type D thinking (shown in Fig. 4) shows multilogical thinking or critical thinking. Multilogical thinking is the opposite of monological thinking – that is thinking that considers opposite points of view, and examines both supporting and opposing arguments for each view considered (Paul, 1995). Only two participants fell into this category, namely, P98 and P105. P98 scored high on the EBQ and P105 low.

Figure 2. Diagram type B thinking

Figure 3. Diagram type C thinking
The results showed that there were participants, i.e., P105, P99, P100, who performed low on the EBQ but well on the problem-solving task, and also there were participants, i.e., P116, P112, and P107, who performed high on the EBQ but poorly on the problem-solving task. Based on these results, it seems that other context-dependent factors affected participants’ performance on the ill-structured issue.

**Paired thinking**

The analysis of the transcripts from the paired sessions is shown in Table 2. For each participant, Table 2 provides information regarding the epistemological beliefs stage (High or Low), evaluation of the performance during individual problem solving (i.e., type A thinking, type B thinking, etc.), number of messages posted by the individual during collaboration in his or her dyad, number of times the individual interacted with the other person in the dyad by replying to him/her, number of times the individual disagreed with the other person in the dyad, and lastly a calculated number which constituted the inter-subjectivity index.

The inter-subjectivity index for each dyad, shown in the last column of Table 2, was calculated by dividing the total number of interactions between the partners with the total number of postings from both partners. For example, the calculated inter-subjectivity index for Dyad 8 was found to be .58 (21/(14+22)=21/36). The calculated value shows the degree of interaction between the two members of the dyad. For example, P116 contributed a total of 14 messages 11 of which were replies to P105. P105 posted a number of 22 messages 10 of which were replies to P116. Thus, Dyad 8 compared to the rest of the dyads had a high degree of interaction (i.e., a high inter-subjectivity index) since the two members highly considered each other in the communication process. It is interesting to point out that as shown in Table 2, the dyads with the largest number of messages posted by both partners, that is, Dyads 3, 6, 9, and 10, had the lowest inter-subjectivity index signifying that group members did not manage to interact with their partners effectively; instead each member posted his/her messages without considering the postings of the other person. On the other hand, Dyad 8 had the smallest number of messages posted but the largest inter-subjectivity index showing the effective interaction between the members of the dyad. The results showed that there were participants, such as P105, P95, P113, and P99, with low epistemological beliefs scores, who performed well on the
collaborative task in terms of intersubjectivity. Also there were participants such as P108, P115, P106, P98, and P104, with high scores on the epistemological beliefs questionnaire who performed poorly on the collaborative task.

Based on these mixed results, personal epistemology did not seem to be related to individuals’ contribution and performance on the collaborative problem-solving context. Furthermore, an individual’s solo performance on the ill-structured problem did not seem to be related with the individual’s performance on the collaborative problem-solving task. A good example of this is Dyad 8. Specifically, the two members of Dyad 8 were P116, who scored high on the EBQ but had the poorest solo performance on the task, and P105 who scored low on the EBQ but had the highest solo task performance. Together, P116 and P105 managed to achieve the highest inter-subjectivity index. The results imply that contextual factors might affect an individual’s performance on an ill-structured problem-solving task more so than their epistemological beliefs scores.

### Table 2. Paired thinking

<table>
<thead>
<tr>
<th>Dyad Member ID</th>
<th>Epistemological Stage</th>
<th>Solo Thinking</th>
<th># of contributions</th>
<th># of interactions</th>
<th># of Disagreements</th>
<th>Inter-subjectivity index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 P107 P113</td>
<td>H Type B</td>
<td>56</td>
<td>17</td>
<td>6</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>2 P102 P114</td>
<td>H Type C</td>
<td>17</td>
<td>8</td>
<td>1</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>3 P108 P115</td>
<td>H Type B</td>
<td>26</td>
<td>6</td>
<td>2</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>4 P100 P109</td>
<td>L Type C</td>
<td>21</td>
<td>6</td>
<td>2</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>5 P99 P111</td>
<td>L Type C</td>
<td>31</td>
<td>7</td>
<td>4</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>6 P106 P110</td>
<td>H Type B</td>
<td>43</td>
<td>11</td>
<td>0</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>7 P112 P95</td>
<td>H Type B</td>
<td>24</td>
<td>9</td>
<td>6</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>8 P116 P105</td>
<td>H Type A</td>
<td>14</td>
<td>11</td>
<td>0</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>9 P98 P103</td>
<td>H Type D</td>
<td>30</td>
<td>11</td>
<td>1</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>10 P104 P96</td>
<td>H Type B</td>
<td>32</td>
<td>12</td>
<td>0</td>
<td>0.30</td>
<td></td>
</tr>
</tbody>
</table>

### Differences between solo and paired transcripts

The analysis of the solo and paired transcripts identified 19 different elements of thinking, and these are shown along with their descriptions in Table 3.

### Table 3. Elements of thinking

<table>
<thead>
<tr>
<th>Elements</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information from reading materials</td>
<td>Inf(M)</td>
<td>Information present in the reading materials provided to learners.</td>
</tr>
<tr>
<td>Cultural Identity</td>
<td>CId</td>
<td>Knowledge that is directly or indirectly related to, and could only be known from, the learner’s culture, as defined by his or her cultural identity.</td>
</tr>
<tr>
<td>Emotion</td>
<td>E</td>
<td>Knowledge, experience, event, or activity that is either directly or indirectly emotionally charged</td>
</tr>
<tr>
<td>Information from personal experience</td>
<td>PE</td>
<td>Knowledge, experience, activity, or event that is derived from the individual’s personal experience,</td>
</tr>
<tr>
<td>Information from other sources</td>
<td>OS</td>
<td>Knowledge, experience, activity, or event that is not derived from the individual’s personal experience,</td>
</tr>
</tbody>
</table>
The 19 elements can be categorized into cognitive, cultural and emotional elements. Cognitive elements are directly related to reasoning, cultural elements are related to one’s culture or cultural identity, and emotional elements are primarily related to the learners’ feelings. As shown in Table 3, the average number of elements when participants thought about the problem alone was 16.39 (SD = 9.92), but when they were put into dyads and were asked to think about the problem with another participant the average number of elements per participant increased dramatically to 33.06 (SD = 13.16). Also solo thinking was more likely to include value judgments not supported by evidence (Mean = 6.39, SD = 6.84), value judgments supported by evidence in the form of information given in the reading materials (Mean = 4.67, SD = 2.52), value judgments supported by evidence in the form of information from other sources (Mean = 2.22, SD = 2.10), information from reading materials (Mean = 1.11, SD = 1.88), inferences (Mean = .50, SD = .79), and value judgments supported by evidence in the form of cultural identity (Mean = .44, SD = .62).

Similarly, the elements of an individual’s thinking when he or she thought about the problem in a dyad was more likely to include value judgments not supported by evidence (Mean = 12.11, SD = 6.00), social acknowledgment (Mean = 4.17, SD = 2.80), questions asking for information (Mean = 3.67, SD = 2.33), value judgments supported by evidence in the form of information from other sources (Mean = 3.22, SD = 2.26), evaluative questions (Mean = 2.39, SD = 2.40), value judgments supported by evidence in the form of information given in the reading materials (Mean = 1.67, SD = 1.57), inferences (Mean = 1.50, SD = 1.50), information from reading materials (Mean = .78, SD = 1.48), value judgments supported by evidence in the form of cultural identity (Mean = .67, SD = .84), and value judgments supported by evidence in the form of an emotion (Mean = .44, SD = .71).
Table 4. Descriptive statistics for the elements of thinking in solo and paired problem-solving contexts

<table>
<thead>
<tr>
<th>Element Code</th>
<th>Solo</th>
<th></th>
<th></th>
<th></th>
<th>Paired</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Mean</td>
<td>SD</td>
<td>Frequency</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inf(M)</td>
<td>20</td>
<td>1.11</td>
<td>1.88</td>
<td>14</td>
<td>.78</td>
<td>1.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cid</td>
<td>1</td>
<td>.06</td>
<td>.24</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>.11</td>
<td>.32</td>
<td>6</td>
<td>.33</td>
<td>.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>2</td>
<td>.11</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>4</td>
<td>.22</td>
<td>.55</td>
<td>12</td>
<td>.67</td>
<td>1.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inference</td>
<td>9</td>
<td>.50</td>
<td>.79</td>
<td>27</td>
<td>1.50</td>
<td>1.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VJ</td>
<td>115</td>
<td>6.39</td>
<td>6.84</td>
<td>218</td>
<td>12.11</td>
<td>6.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VJ(M)</td>
<td>84</td>
<td>4.67</td>
<td>2.52</td>
<td>30</td>
<td>1.67</td>
<td>1.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VJ(Cid)</td>
<td>8</td>
<td>.44</td>
<td>.62</td>
<td>12</td>
<td>.67</td>
<td>.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VJ(E)</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>8</td>
<td>.44</td>
<td>.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VJ(PE)</td>
<td>1</td>
<td>.06</td>
<td>.24</td>
<td>1</td>
<td>.06</td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VJ(OS)</td>
<td>40</td>
<td>2.22</td>
<td>2.10</td>
<td>58</td>
<td>3.22</td>
<td>2.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q(I)</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>66</td>
<td>3.67</td>
<td>2.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q(E)</td>
<td>6</td>
<td>.33</td>
<td>.97</td>
<td>43</td>
<td>2.39</td>
<td>2.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q(H)</td>
<td>3</td>
<td>.17</td>
<td>.51</td>
<td>1</td>
<td>.06</td>
<td>.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q(Cl)</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>7</td>
<td>.39</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td>1</td>
<td>.06</td>
<td>.24</td>
<td>75</td>
<td>4.17</td>
<td>2.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>1</td>
<td>.06</td>
<td>.24</td>
<td>7</td>
<td>.39</td>
<td>.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarification</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>6</td>
<td>.33</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>295</td>
<td>16.39</td>
<td>9.92</td>
<td>595</td>
<td>33.06</td>
<td>13.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Repeated measures analyses of variance were subsequently conducted to detect any significant differences between the number of elements of participants’ reasoning when thinking alone and in a dyad. According to the analyses, significant within-subject effects were found for cognitive and emotional elements; that is, Inference ($F = 5.23, p < .05$), Value judgments not supported by evidence ($F = 25.32, p < .01$), Value judgments supported by evidence in the form of information given in the reading materials ($F = 12.79, p < .01$), Value judgments supported by evidence in the form of an emotion ($F = 8.00, p < .05$), Evaluative questions ($F = 13.83, p < .01$), Social acknowledgment ($F = 119.04, p < .01$), and Clarification ($F = 9.00, p < .05$). The analyses did not reveal any significant between-subject effects.

Discussion

The paper reports on the results of a mixed-method exploratory study that sought to better understand how participants with different personal epistemological beliefs reasoned about an ill-structured issue individually and then with others in dyads. According to the qualitative results of this study there was not a systematic connection between epistemological beliefs and ill-structured problem solving in either solo or paired contexts. For example, there were participants who scored low on the epistemological beliefs questionnaire but achieved high individual performance on the ill-structured problem, and participants who scored high on the epistemological beliefs test and achieved low individual problem-solving performance. Similarly, participants with low epistemological beliefs scores achieved high group performance, and participants with high epistemological beliefs scores achieved low group performance. There were also instances where participants with high epistemological beliefs scores achieved high individual problem-solving performance but very low group problem-solving performance. Thus, according to the results of this study, it seems that ill-structured problem solving entails some unique characteristics that influence one’s reasoning about the problem.

The results from the analysis of the transcripts showed fundamental differences between solo and paired thinking. One difference was that while participants when thought about the problem alone did not always reason within a well-supported point of view, all participants in their dyads clearly reasoned within a point of view. This of course can be explained by the fact that participants in their dyads explicitly asked each other about their point of view before they began the discussion. Another major difference between individual and paired thinking was that when
participants thought about the problem alone, they used mostly the materials we gave them in order to support their point of view, while, in their dyads, they used mostly emotional and cultural statements to explain their reasoning. Thus, the fact that the problem was ill-structured, highly controversial, and emotional, influenced the way participants thought about it. Therefore, another difference between solo and paired thinking about an ill-structured controversial issue was that individual thinking was mostly cognitive, whereas social thinking entailed strong cultural and emotional elements. Thus, the results indicate that problem solving within a social context may trigger more emotional activity for an individual than when he or she thinks alone. Furthermore, in this study, the ill-structured problem that was given to the participants was also highly emotional and culturally-based. Based on the results, it seems that the relationship between personal epistemology and problem solving can be better understood if it is conducted in a way, so that the intricacies of the specific context are considered carefully.

All things considered, this is an exploratory study with one small localized sample, thus obviously no generalizations can be drawn at this point. It will be valuable though if future studies with larger samples further examine the issues discussed herein, so that ultimately a theory about social epistemology can be derived. In addition, this study only examined one type of problem, namely a highly controversial and emotional ill-structured problem, thus future experimental designs with different kinds of problems can provide valuable insights about the extent to which different problem types require different sets of epistemological beliefs, as well as different cognitive, metacognitive, and affective skills.

Concluding remark

Based on the results of this small exploratory study, it has become evident that in order to better understand the complex relationship between epistemological beliefs and reasoning, a departure from cold cognition toward more integrative approaches is necessary in order to understand the contextual and dynamic nature of intellectual functioning in partnership with epistemological development (Labouvie-Vief, 1990; Pintrich, Marx, & Boyle, 1993; Sinatra, 2005). We consider this study to be an important first step toward developing a theory about social epistemology and social problem solving. The results of the study indicated that thinking is not a decontextualized construct, but a construct that develops in a context where emotions, culture, and social experiences, seem to be inextricably related in intellectual functioning and development (Brackett, Lopes, Ivcevic, Mayer, & Salovey, 2004; Li & Fischer, 2004). Future studies toward examining these issues with larger samples and systematic experimental designs will be of utmost importance to the research community.

References


A Study on Exploiting Commercial Digital Games into School Context

Hercules Panoutsopoulos and Demetrios G. Sampson*
Department of Digital Systems, University of Piraeus & Doukas School, Greece // *Department of Digital Systems, University of Piraeus & Informatics and Telematics Institute, Center for Research and Technology – Hellas, Greece
herculespanoutsopoulos@gmail.com // sampson@iti.gr

ABSTRACT
Digital game-based learning is a research field within the context of technology-enhanced learning that has attracted significant research interest. Commercial off-the-shelf digital games have the potential to provide concrete learning experiences and allow for drawing links between abstract concepts and real-world situations. The aim of this paper is to provide evidence for the effect of a general-purpose commercial digital game (namely, the “Sims 2-Open for Business”) on the achievement of standard curriculum Mathematics educational objectives as well as general educational objectives as defined by standard taxonomies. Furthermore, students’ opinions about their participation in the proposed game-supported educational scenario and potential changes in their attitudes toward math teaching and learning in junior high school are investigated. The results of the conducted research showed that: (i) students engaged in the game-supported educational activities achieved the same results with those who did not, with regard to the subject matter educational objectives, (ii) digital game-supported educational activities resulted in better achievement of the general educational objectives, and (iii) no significant differences were observed with regard to students’ attitudes towards math teaching and learning.

Keywords
Commercial off-the-shelf games, Game-supported educational activities, School math teaching and learning

Introduction
Digital game-based learning is a research field within the wider context of technology-enhanced learning that has attracted, during the last few years, the interest of both the research and educational community (Kirriemuir & McFarlane, 2004; Sandford & Williamson, 2005; Sandford et al., 2006; Van Eck, 2007, Chen & Chan, 2010). Connolly and Stansfield (2007) define digital game-based learning as “the use of a computer games-based approach to deliver, support and enhance teaching, learning, assessment, and evaluation”, whereas Prensky (2007, pp. 145–146) stresses the additional educational value of digital game-based learning by defining it as an approach based on the integration of educational content into digital games and leading to the achievement of the same or better results, in comparison to traditional instructional approaches. Furthermore, Chen and Wang (2009) focus on the motivational aspect of digital games and their potential to facilitate active construction of knowledge by defining digital game-based learning as “an effective means to enable learners to construct knowledge by playing, maintain higher motivation and apply acquired knowledge to solve real-life problems”.

Research interest regarding digital game-based learning can be first of all attributed to the fact that digital games engage and motivate people of all ages (Sautler, 2007). Furthermore, by simulating real-world situations (Winn, 2002) and presenting ill-defined problems (Klopfer, 2008, p. 17; Whitton, 2010, p. 51), general-purpose commercial games bare the potential to situate players’ activities within authentic and meaningful contexts (Prensky, 2007, p. 159; Gee, 2007, pp. 71–110; Whitton, 2010, p. 46) and offer opportunities for learning by applying trial-and-error approaches (Oblimger, 2004; Prensky, 2007, pp. 158–159; Chen & Shen, 2010). Players are able to engage in active explorations, formulate and test hypotheses within the virtual world of the game, and, based on feedback, confirm or reject them (Gee, 2007, p. 105).

The engagement and motivation that games offer alongside with their potential to provide concrete learning experiences has attracted significant research interest with regard to the integration of commercial games into formal educational settings as well as the development and use of specially-designed educational games (Kirriemuir & McFarlane, 2004; Sandford & Williamson, 2005; Van Eck, 2006). While there is a large number of research studies considering the use of educational digital games for delivering educational content (e.g. Rosas et al., 2003; Williamson Shaffer, 2006; Bottino et al., 2007; Ke, 2008; Sisler & Brom, 2008; Lim, 2008; Annetta et al., 2009; Papastergiou 2009; Tuzun et al., 2009), there are relatively few studies investigating methods of integrating
commercial off-the-shelf digital games into existing teaching practices (e.g. Squire & Barab, 2004; Egenfeldt-Nielsen, 2005; Sandford et al., 2006; Robertson & Miller, 2009; Tanes & Cemalcilar, 2009).

In this context, the aim of this paper is to provide evidence for the effect that commercial simulation games can have on the achievement of standard curricula educational objectives when used as part of wider sets of appropriately designed educational activities. More specifically, our work focuses on investigating the influence of a commercial business simulation game (namely, the “Sims 2 – Open for Business”) on achieving educational objectives related to the subject matter of Mathematics as well as general educational objectives defined by standard taxonomies. Furthermore, students’ opinions about the use of the selected digital game and potential changes in their attitudes toward math teaching and learning in junior high school are investigated.

**Literature review**

Digital game-based learning research investigates, among others, methods of integrating digital games (commercial or educational) into existing teaching practices with the purpose to facilitate the achievement of standard curricula educational objectives, increase students’ motivation, and develop positive attitudes toward specific subjects and/or school education in general (e.g. Rosas et al., 2003; Squire & Barab, 2004; Egenfeldt-Nielsen, 2005; Williamson Shaffer, 2006; Bottino et al., 2007; Ke, 2008; Robertson and Miller, 2009; Papastergiou 2009; Tuzun et al., 2009). In particular for Mathematics teaching and learning at school level there are a number of studies mainly focusing on the implementation and evaluation of educational designs aiming at the achievement of subject matter educational objectives with the support of specially-designed educational games (Rosas et al., 2003; Williamson Shaffer, 2006; Bottino et al., 2007; Ke, 2008).

Evidence provided from research shows that using educational games as part of Mathematics teaching at school level can be at least as effective as non-gaming approaches with regard to the achievement of subject matter educational objectives (Rosas et al., 2003; Williamson Shaffer, 2006; Ke, 2008). By engaging students in long-lasting game-supported educational activities there is potential for enhancing the development of problem-solving skills and achieving improved results in mathematics exams (Bottino et al., 2007). With regard to the need for supporting students draw links between school-based mathematics and real-world situations (Lowrie, 2005), Williamson Shaffer (2006) shows that using role-playing educational games for designing and implementing meaningful activities allows for providing students with concrete examples highlighting potential uses of abstract mathematical concepts and procedures in specific domains. In this context, Ke (2008) stresses the need for appropriate educational designs targeting at framing the use of educational games by claiming that monitoring activities with games and supporting them by supplementary tools and/or resources are necessary for the achievement of intended learning outcomes. To this end, she provides evidence indicating that students do not use feedback provided from games in order to reflect on their actions and hence lack opportunities for constructing and evaluating new knowledge.

Continuing with the effects that innovations based on the use of digital games can have on school math education, most studies demonstrate significant increase in students’ motivation as well as their interest toward the subject matter of Mathematics and/or school education in general (Rosas et al., 2003; Williamson Shaffer, 2006; Ke, 2008; Lim, 2008; Robertson & Miller, 2009). Important issues that have been highlighted are the improvement of relationships between students (Robertson & Miller, 2009) as well as improvement of communication and collaboration between students and teachers (Rosas et al., 2003). Positive effects have also been noticed with regard to students’ discipline, on task concentration, peer collaboration, perseverance in task completion (Rosas et al., 2003), and responsibility (Rosas et al., 2003; Robertson & Miller, 2009).

Finally, there is a small number of studies targeting at providing evidence for the impact of game-supported educational innovations on the development of Mathematics related skills and competencies (e.g. Bottino et al., 2007; Robertson & Miller, 2009). More specifically, Robertson and Miller (2009) present research findings showing positive effects of puzzle games on elementary school students’ mental computational skills such as accuracy and speed in conducting numerical operations, whereas Bottino et al. (2007) claim that appropriate educational designs, supported by the use of educational games, can promote the development of critical thinking skills by engaging students in formulation and testing of hypotheses, reflection activities, and drawing inferences.
As evidenced by the literature review there is a significant number of research studies focusing on the effects that specially-designed educational games can have when used either in the context of school math education (e.g. Rosas et al., 2003; Williamson Shaffer, 2006; Bottino et al., 2007; Ke, 2008) or as part of teaching subjects other than Mathematics (e.g. Papastergiou, 2009; Annetta, 2009; Tuzun et al., 2009). On the other hand, there are relatively few studies investigating the potential use of general-purpose commercial games in the context of school-based education in general (e.g. Squire & Barab, 2004; Egenfeldt-Nielsen, 2005; Sandford et al., 2006; Tanes & Cemalcilar, 2009) and even less with regard to Mathematics teaching and learning at school level in particular (e.g. Robertson & Miller, 2009). Thus, the main purpose of our study is to investigate methods of integrating commercial off-the-shelf digital games, and more specifically simulation games, into the context of Mathematics teaching by proposing and implementing an appropriately designed scenario of game-supported educational activities and providing evidence for their effect on achieving standard curriculum educational objectives.

Design and implementation of research

Research questions

Based on the literature review, we propose the following questions to be researched for the purpose of our study:

- **RQ1**: Is the proposed educational design, based on the use of the commercial business simulation game “Sims 2 – Open for Business”, more effective than a non-gaming approach in terms of achieving standard curriculum mathematics educational objectives?

- **RQ2**: Is the proposed educational design, based on the use of the commercial business simulation game “Sims 2 – Open for Business”, more effective than a non-gaming approach in terms of achieving general educational objectives, as defined by standard taxonomies?

- **RQ3**: What are students’ opinions about the use of the game “Sims 2-Open for Business” in the context of Mathematics teaching and do their attitudes toward school math teaching and learning change after having participated in the proposed game-supported educational activities?

Research method and study participants

The method that was employed for researching the aforementioned questions was field experiment with one experimental and one control group and the assignment of a post-test (Cohen, Manion & Morrison, 2008, p. 278). Field experiment is a variation of the experimental method, commonly used in cases of empirical studies conducted in educational settings (Cohen, Manion & Morrison, 2008, p. 274). It allows for investigating potential effects of educational innovations (often in comparison to other mainstream practices) as well as observing interactions taking place in natural settings and hence it is considered as appropriate for the purpose of our study.

Our study participants were 59 students (N = 59), at the age of 13–14 years old, attending the second grade of a private junior high school located in Athens, Greece. Students belonged to two different classes (classes A and B), one of which (class A) was the experimental group (number of students = 30) whereas the other (class B) was the control group (number of students = 29).

Research instruments

Questionnaires

Background questionnaires and post-questionnaires were used in the beginning and at the end of our research respectively in order to gather data for shaping students’ profile and investigate potential changes in their attitudes toward school math teaching and learning. The background questionnaire consisted of three parts and a total number of 31 questions with the first two parts including thirteen Likert type questions regarding attitudes toward the use and usefulness of computers in the educational process (Texas Center for Educational Technology, 2010) and eight questions regarding students’ involvement in gaming activity (Pew Internet & American Life Project, 2010) respectively. The third part included 10 Likert type questions targeting at investigating attitudes toward school math teaching and learning (Kislenko et al., 2005) in the beginning of our research.
The post-questionnaire was used after the implementation of the game-supported educational scenario. It consisted of two parts with its first part being the same with the third part of the background questionnaire and its second part including the following two open-ended questions:

- Q1: What is your opinion about the use of the game in the context of Mathematics teaching?
- Q2: Do you believe that the use of the game helped you, in any way, to understand better the mathematical concepts that were taught?

Post-test

Tests are research instruments which, in the context of digital game-based learning research are most commonly used for the assessment of subject matter educational objectives (e.g. Rosas et al., 2003; Ke, 2008; Egenfeldt-Nielsen, 2005; Papastergiou, 2009). For the purpose of our study, a post-test targeting at the assessment of subject matter educational objectives was assigned to students of both groups. It contained matching pairs of items, true/false statements, as well as two open-ended questions, and its design was based on proposed good practice standards (Cohen, Manion & Morrison, 2008, pp. 426–429).

Worksheets

In game-supported educational designs, learners’ activities with games are often supplemented by tools such as worksheets (Sandford et al., 2006; Ke, 2008) which are used with the aim to facilitate necessary reflection activities. In the context of our research, worksheets, designed by researchers, were used by students in order to formulate hypotheses, write down the results of their hypotheses’ testing, and provide explanations for observed results. This instrument was used for gathering data and providing evidence for the effect of the proposed educational design on the achievement of general educational objectives.

Selection of digital game and pedagogical framework

The game that was selected to support the proposed educational activities was “Sims 2–Open for Business”, a commercial business simulation game which engages players in activities requiring data monitoring, strategic thinking, decision making, as well as planning and performing actions related to managing a business and keeping customers satisfied. The game allows players to set price of products, hire employees based on specific criteria and assign tasks to them by taking into consideration their talents and interests. As a simulation game it depicts a simplified version of reality (Herz, 1997, pp. 215–223). Sophisticated graphics and advanced sound effects help to create a rich and interactive environment in which players have a sense of control (Herz, 1997, pp. 215–223) and are offered opportunities to get engaged in active explorations, hypotheses testing, and discovery of causal relationships between game variables.

Exploiting digital games for educational purposes requires careful consideration of a number of issues that can ensure the alignment of game features with the intended learning outcomes. Thus, selecting an appropriate pedagogical approach for framing the game-supported educational activities is considered as highly important. The pedagogical approach that was employed in the context of our study was the problem-solving model (Eggen & Kauchak, 2006, pp. 252–259).

Problem-based learning involves the assignment of ill-defined, real-world problems to students (Whitton, 2010, p. 50) who are prompted to collaborate in order to design, implement, and evaluate strategies for solving them (Eggen & Kauchak, 2006, p. 250). Educational designs based on the problem-solving model allow for engagement in authentic and meaningful activities in the context of which learners are able to draw links between abstract concepts and real-world practices, as well as, to develop skills that can be further applied to other contexts (Eggen & Kauchak, 2004). Furthermore, games and simulations are considered as digital tools, commonly employed by instructors when developing educational designs based on the problem-solving paradigm (de Freitas, 2006).
Game-supported educational design

The design of both study groups’ educational scenarios was based on a common pedagogical approach, namely the problem-solving model. The intended educational objectives, as well as, the activities that were designed to facilitate their achievement are described in the following two sections.

Educational objectives

In our experiment, the subject matter educational objectives, explicitly described by the Greek National Curriculum (2003), refer to linear functions (namely “y = ax” and “y = ax + b”) and they relate to: (i) drawing the graphs of linear functions on a set of cartesian axes, (ii) finding the slope of a line when the algebraic type of the corresponding linear function is provided, (iii) finding the points of intersection between the graph of a linear function and the two axes, and (iv) finding the algebraic type of a linear function when specific data are given (e.g. the slope of the line and a point on the graph).

General educational objectives are aligned with the upper levels of the cognitive domain of Bloom’s taxonomy (namely “analysis”, “synthesis”, and “evaluation”) and can be achieved by designing and implementing educational activities targeting at involving students in actions like: (i) comparing and contrasting, (ii) explaining reasons for, and (iii) evaluating results (Falconer et al., 2006).

Scenarios of educational activities

The problem-solving model consists of five phases of educational activities (Eggen & Kauchak, 2006, pp. 252–259) as shown in Figure 1.

![Figure 1. Phases of educational activities of the problem-solving model](image)

However, when integrating digital games into the educational process it is important to design and implement appropriate activities targeting at familiarizing students with the selected game (Sandford et al., 2006; Whitton, 2010, p. 82). In the case of the experimental group’s educational scenario, an additional phase of activities was inserted between phases one and two of the problem-solving model as shown in Figure 2.

![Figure 2. Phases of educational activities of the experimental group’s scenario](image)

Activities of the experimental group’s educational scenario were described by adopting the DialogPlus taxonomy of educational activities and are summarized in Table 1 below. As indicated by the brief description of activities in Table 1, students of the experimental group (class A) were assigned a problem targeting at investigating issues related to the management of an enterprise. For testing alternative solutions to the given problem, students were divided into six groups and two sessions of educational activities (of two didactic hours each), fully supported by the selected game, were implemented.
Table 1. Phases and activities of the experimental group’s scenario

<table>
<thead>
<tr>
<th>Phase</th>
<th>Implemented Educational Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1: Identify the problem</td>
<td>The teacher:</td>
</tr>
<tr>
<td></td>
<td>• makes a brief description of the game-supported educational activities that will be implemented.</td>
</tr>
<tr>
<td></td>
<td>• presents the intended educational objectives,</td>
</tr>
<tr>
<td></td>
<td>• presents the problem to be solved.</td>
</tr>
<tr>
<td></td>
<td>Students interact with the game in order to familiarize themselves with the interface and the actions that can be performed.</td>
</tr>
<tr>
<td></td>
<td>A class-based discussion takes place where students express their opinions with regard to issues related to the given problem’s solution.</td>
</tr>
<tr>
<td></td>
<td>The teacher with the support of students constructs a mind map depicting relations between these issues.</td>
</tr>
<tr>
<td>Phase 2: Familiarizing students with the game</td>
<td>The teacher:</td>
</tr>
<tr>
<td></td>
<td>• makes a brief presentation with regard to the content and objectives of the game,</td>
</tr>
<tr>
<td></td>
<td>• performs a live, in class, demonstration of the game.</td>
</tr>
<tr>
<td></td>
<td>Students interact with the game in order to familiarize themselves with the interface and the actions that can be performed.</td>
</tr>
<tr>
<td></td>
<td>A class-based discussion takes place with regard to actions that students should perform, with the support of the selected game, in order to test potential solutions to the problem.</td>
</tr>
<tr>
<td>Phase 3: Represent the problem</td>
<td>The teacher with the support of students constructs a mind map depicting relations between these issues.</td>
</tr>
<tr>
<td></td>
<td>A class-based discussion takes place with regard to issues related to the given problem’s solution.</td>
</tr>
<tr>
<td>Phase 4: Select a strategy</td>
<td>Students:</td>
</tr>
<tr>
<td></td>
<td>• collaborate in order to test solutions within the virtual world of the game,</td>
</tr>
<tr>
<td></td>
<td>• work out arithmetic examples in order to investigate causal relationships between specific game variables and try to derive the underlying algebraic formulas.</td>
</tr>
<tr>
<td>Phase 5: Implement the strategy</td>
<td>The teacher:</td>
</tr>
<tr>
<td></td>
<td>• monitors students’ activities with the game and provides support with regard to the implementation of the agreed plan of actions,</td>
</tr>
<tr>
<td></td>
<td>• presents new mathematical concepts related to linear functions.</td>
</tr>
<tr>
<td>Phase 6: Evaluate results</td>
<td>Students:</td>
</tr>
<tr>
<td></td>
<td>• collaborate in order to develop their final proposals-solutions to the given problem,</td>
</tr>
<tr>
<td></td>
<td>• present their final proposals.</td>
</tr>
</tbody>
</table>

Members of each group were prompted to select a virtual enterprise and investigate effects of actions that the game allows for (e.g. hiring employees and assigning tasks to them, setting prices for products, increasing employees’ salaries etc) on the status of their business. To this end, they were asked to formulate hypotheses, test them within the virtual world of the game, confirm or reject these hypotheses, provide explanations for observed results, and develop final proposals-solutions to the given problem. As part of formulating their hypotheses, students were expected to explicitly describe actions to be performed, with regard to their virtual enterprise’s management, and anticipated results. After having applied the proposed actions, students used feedback provided from the game in order to compare the status of their business before and after the testing of hypotheses and hence confirm or reject them. Figure 3 illustrates the type of feedback that the game provides with regard to the virtual enterprise’s status.

![Figure 3. Feedback regarding the status of virtual enterprise](image-url)
Providing explanations for observed results allowed for reflecting upon performed actions as well as discovering causal relationships between performed actions and their outcomes. The development of final proposals was the result of the evaluation of actions performed within the virtual world of the game and their effects on the status of the virtual enterprise. Figure 4 shows the actions that students performed with the support of the selected game.

![Figure 4. Actions performed with the support of game](image)

With regard to subject matter educational objectives, students of the experimental group were asked to work out arithmetic examples which would help them derive algebraic formulas that highlight relationships between specific game variables (e.g. “wholesale cost of a product” and “retail cost of a product”). These activities served as a starting point for the presentation of new mathematical concepts by the teacher (related to linear functions) and students were provided with concrete examples targeting at helping them draw links between abstract mathematical concepts and variables of the game. Figure 5 illustrates the framework that was adopted for the design and implementation of the experimental group’s game-supported educational scenario.

![Figure 5. Framework for the design of game-supported activities](image)

The control group’s scenario of educational activities was also based on the problem-solving model with students being presented with a problem similar to that of the experimental group’s. More specifically, students of the control group (class B) were assigned the role of a computer store’s sales manager and prompted to collaborate in order to develop a proposal for a potential customer. By considering issues such as specifications imposed by the client, cost of material, salaries and expertise of employees, profit percentage, and time needed to satisfy the customer’s request, they were asked to develop alternative solutions to the given problem and finally propose the one that would best meet the aforementioned criteria. For the investigation of alternative solutions to the given problem, students were provided a period of time equal to that of the experimental group. Necessary data could be extracted from websites and printed material provided by the teacher.
Research results

Describing students’ profile

Participating students were at the age of 13–14 years old when our study was conducted and they are coming from families with an average or high socio-economic background. As evidenced by data gathered from background questionnaires, our research subjects were familiar with the use of computers, which constituted an integral part of their everyday life and culture, and reported that they are convinced that digital technology can have a positive effect on the achievement of educational objectives. With regard to their involvement in digital gaming activity, 98.2% of the total sample reported playing digital games, with the frequency of the gaming activity ranging from many times a day to a few times a month (94.5% of the total sample) and its duration from 1 to 4–5 hours per time (96.3% of the total sample).

As far as students’ background mathematics knowledge is concerned, final grades at the end of the previous school year were taken into account and the two groups’ mean scores were compared. The mean score of the experimental group was 79.46% (SD = 11.574), whereas the mean score of the control group was 78.97% (SD = 12.563). The t-test (Cohen, Manion & Morrison, 2008, pp. 543–546) that was conducted revealed no significant differences between the two study groups’ mean scores (t = 0.156, df = 55, two-tailed, p = 0.877).

Effect on the achievement of subject matter educational objectives

The effect of the proposed game-supported educational design on the achievement of standard curriculum mathematics educational objectives was measured by conducting the t-test in order to compare the mean scores of the two groups’ post-tests. Results are analytically presented in Table 2.

<table>
<thead>
<tr>
<th>Groups’ scores</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group</td>
<td>29</td>
<td>60.52%</td>
<td>19.335</td>
<td>3.590</td>
</tr>
<tr>
<td>Control Group</td>
<td>28</td>
<td>57.86%</td>
<td>27.770</td>
<td>5.248</td>
</tr>
</tbody>
</table>

As far as the comparison of the two groups’ mean scores is concerned, Levene’s test showed that no equal variances could be assumed (p = 0.006 < 0.05) and the results of the t-test corresponding to this case, which are analytically presented in Table 3, revealed no significant differences between the two study groups’ mean scores (t = 0.418, df = 48.041, two-tailed, p = 0.678).

<table>
<thead>
<tr>
<th>Groups’ scores</th>
<th>Levene’s Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>8.309</td>
<td>.006</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>.418</td>
<td>48.041</td>
</tr>
</tbody>
</table>

Effect on the achievement of general educational objectives

Monitoring students’ activities with games in order to ensure the achievement of the intended educational objectives (Torrente et al., 2009, pp. 1–18) as well as developing specific criteria for assessing students’ performance in the
context of problem-based educational scenarios (Eggen & Kauchak, 2006, pp. 273–276) are considered as highly important. To this end, we used appropriately designed worksheets for supporting educational activities with the selected game and employed specific assessment criteria, fully aligned with the general educational objectives. Assessment criteria as well as their alignment with the game-supported educational activities and the intended educational objectives are presented in Table 4.

<table>
<thead>
<tr>
<th>Table 4. Criteria for assessing the experimental group’s worksheets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digital game-supported activities</strong></td>
</tr>
<tr>
<td>Hypotheses formulation</td>
</tr>
<tr>
<td>Comparison of data provided from game menus before and after the testing of hypotheses</td>
</tr>
<tr>
<td>Justification of results</td>
</tr>
</tbody>
</table>

Results from the assessment of the game-supported educational activities are presented in Table 5. Results are presented for each one of the six groups that students of the experimental group formed and for each one of the two sessions of game-supported activities.

<table>
<thead>
<tr>
<th>Table 5. Results from the assessment of the experimental group’s worksheets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Groups of students</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>criterion 1</td>
</tr>
<tr>
<td>criterion 2</td>
</tr>
<tr>
<td>criterion 3</td>
</tr>
<tr>
<td>criterion 4</td>
</tr>
</tbody>
</table>

Data presented in Table 5 show that most groups’ activities fulfilled the employed criteria as well as an improvement in performance during the second session. Furthermore, the final solutions that students of the experimental group proposed indicated successful engagement in the evaluation of performed actions’ outcomes and hence positive effect of the game-supported educational activities on the achievement of the related objective (namely “evaluating results”).

As far as the control group is concerned, students’ activities were assessed by employing equivalent criteria fully aligned with the intended educational objectives. Assessment criteria as well as their alignment with performed activities and intended educational objectives are displayed in Table 6.

<table>
<thead>
<tr>
<th>Table 6. Criteria for assessing the control group’s worksheets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Educational activities</strong></td>
</tr>
<tr>
<td>Comparison of alternative solutions to the given problem</td>
</tr>
<tr>
<td>Development of final proposal-solution to the given problem</td>
</tr>
</tbody>
</table>
Assessment results for each one of the six groups that students of the control group formed showed that only one of them managed to develop more than one alternative solution to the given problem. As a consequence, most of the control group’s students did not manage to engage in actions requiring comparison and contrasting of alternative solutions (criterion 1) and consequently actions requiring justification (criterion 2) and evaluation of results (criterion 3). Thus, achievement of the intended general educational objectives cannot be inferred in this case.

Students’ opinions about the use of the game and investigation of changes in attitudes toward school math teaching and learning

Students’ opinions about the use of the business simulation game “Sims 2-Open for Business” were investigated by the assignment of two open ended questions (they are analytically presented in the game-supported educational design section) after the implementation of the proposed educational scenario. As evidenced by the analysis of answers that were provided to the first question (namely “What is your opinion about the use of the game in the context of Mathematics teaching?”), students reported that the implementation of the game-supported educational activities was pleasant and innovative, attracted their interest, and provided opportunities for investigating and understanding real-world situations. Furthermore, there were answers highlighting the proposed educational design’s effect on understanding mathematical concepts as well as the limited duration of the implemented activities. The main issues that were revealed from students’ answers as well as their frequencies are presented in Table 7 below.

<table>
<thead>
<tr>
<th>Issues highlighted by students’ answers</th>
<th>Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interesting and innovative approach to the lesson</td>
<td>55.2 %</td>
</tr>
<tr>
<td>Effect on understanding the mathematical concepts that were taught</td>
<td>51.7 %</td>
</tr>
<tr>
<td>A pleasant way to make the lesson</td>
<td>27.6 %</td>
</tr>
<tr>
<td>Opportunities for investigating and understanding real-world situations</td>
<td>20.7 %</td>
</tr>
<tr>
<td>Time constraints</td>
<td>17.2 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 8. Answers provided to the second of the two open-ended questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students’ answers</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>The use of the selected digital game helped me understand the mathematical concepts that were taught.</td>
</tr>
<tr>
<td>The use of the selected digital game helped me partially understand the mathematical concepts that were taught.</td>
</tr>
<tr>
<td>The use of the selected digital game did not help me understand the mathematical concepts that were taught.</td>
</tr>
</tbody>
</table>

With regard to the answers that were provided to the second question (namely “Do you believe that the use of the game helped you, in any way, understand better the mathematical concepts that were taught?”), 55.2% of participating students reported that their involvement in the game-supported educational activities had a positive effect on understanding the mathematical concepts that were taught, whereas 44.8% of the experimental group’s students reported no positive effect of the game. The answers that students provided as well as their frequencies are analytically presented in Table 8 above.

The effect of the game-supported educational scenario on students’ attitudes toward school math teaching and learning was measured by comparing their replies to the 10 Likert type questions included in the third part of the background questionnaire and the first part of the post-questionnaire. The comparison was conducted by employing the Wilcoxon test (Cohen, Manion & Morrison, 2008, pp. 552–554) which results are presented in Table 9 below.

Data presented in Table 9 above show that no significant differences were found in the replies that students provided to 8 out of the 10 questions before and after the implementation of the educational activities. Thus, no significant changes in attitudes toward school math teaching and learning were observed.
Table 9. Results of the Wilcoxon test

<table>
<thead>
<tr>
<th>Statement</th>
<th>Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Statistically significant difference?</th>
</tr>
</thead>
<tbody>
<tr>
<td>The way the subject of Mathematics is taught is interesting.</td>
<td>-.309a</td>
<td>.757</td>
<td>NO</td>
</tr>
<tr>
<td>The way the subject of Mathematics is taught helps me understand the concepts which are taught.</td>
<td>-.598a</td>
<td>.550</td>
<td>NO</td>
</tr>
<tr>
<td>The way the subject of Mathematics is taught helps me understand its usefulness.</td>
<td>-2.306a</td>
<td>.021</td>
<td>YES</td>
</tr>
<tr>
<td>The subject of Mathematics is useful for me in my life.</td>
<td>-.922a</td>
<td>.357</td>
<td>NO</td>
</tr>
<tr>
<td>Mathematics helps me understand life in general.</td>
<td>-1.844a</td>
<td>.065</td>
<td>NO</td>
</tr>
<tr>
<td>Mathematics can help me make important decisions.</td>
<td>-.124a</td>
<td>.901</td>
<td>NO</td>
</tr>
<tr>
<td>Good mathematics knowledge makes it easier to learn other subjects.</td>
<td>-.291a</td>
<td>.771</td>
<td>NO</td>
</tr>
<tr>
<td>The subject of Mathematics is important.</td>
<td>-.741a</td>
<td>.458</td>
<td>NO</td>
</tr>
<tr>
<td>It is important for someone to be good at Maths in school.</td>
<td>-2.211a</td>
<td>.027</td>
<td>YES</td>
</tr>
<tr>
<td>The subject of Mathematics is boring.</td>
<td>-1.006a</td>
<td>.314</td>
<td>NO</td>
</tr>
</tbody>
</table>

a. Based on positive ranks.

Conclusions – Discussion

As evidenced by the analysis of research data, the use of the selected game in the context of an appropriate educational design facilitated the achievement of general educational objectives and was equally effective with the non-gaming approach in terms of achieving standard curriculum mathematics educational objectives. The fact that there are research findings showing that educational games can be as effective as non-gaming approaches, with regard to the achievement of Mathematics related objectives, (e.g. Rosas et al., 2003; Ke, 2008) allows us to infer that not only specially-designed educational games but also general-purpose commercial games can contribute to the achievement of standard curriculum mathematics educational objectives when used as part of appropriately designed activities.

By designing and implementing meaningful activities with the support of the selected game we offered opportunities for engaging students in problem-solving actions. Students were able to formulate and test their own hypotheses, observe the outcomes of their actions, compare and contrast data available from the game, justify and evaluate outcomes of performed actions. Feedback provided from the game as well as its potential to simulate unexpected events were specific features that informed students’ actions within the game world. Supporting game-based activities with appropriately designed worksheets provided the necessary structure and allowed for reflection. As evidenced by the results of our research, students of the experimental group outperformed their control group counterparts with regard to achieving general educational objectives. Thus, commercial simulation games, as opposed to educational games, can be considered as highly interactive environments providing learners with structure and authentic learning contexts. With the support of our findings we can confirm statements highlighting the contribution of commercial off-the-shelf digital games to the achievement of educational objectives aligned with the upper levels of standard taxonomies (Van Eck, 2006).

Finally, participating students commented on the innovative character of the game-based scenario and reported positive effects on understanding real-world situations. However, the limited duration of the proposed educational design did not probably allow for the establishment of intended links between abstract mathematical concepts and real-world situations, at least not to the degree that was expected. Furthermore, expectations that students were likely to have from such an innovation, especially if we consider their gaming experience, can provide explanations for the fact that their attitudes toward school math teaching and learning did not overall change. On the other hand, it must be noticed that digital gaming is generally considered as a leisure activity with no potential implications for learning (Rieber, 1996) and thus the effectiveness of digital game-based learning should be evidenced by further research. To this end, larger scale and longer term research is proposed with an emphasis on the design and implementation of activities highlighting links between school-based mathematics and real-world situations and allowing for interdisciplinary approaches.
References


Torrente, J. et al. (2009). ‘Coordinating Heterogeneous Game-Based Learning Approaches in Online Learning Environments’, in Z.P. Adrian, D. Cheok, & W. Muller (Eds.), *Transactions on Edutainment II, Volume 2*, Berlin, Germany: Springer-Verlag.


Aberrance Detection Powers of the BW and Person-Fit Indices

Tsai-Wei Huang
Department of Counseling, National Chiayi University, 85 Wunlong Village, Minsyong, Chaiyi 62103, Taiwan // twhuang@mail.ncyu.edu.tw

ABSTRACT
The study compared the aberrance detection powers of the BW person-fit indices with other group-based indices (SCI, MCI, NCI, and Wc&Bs) and item response theory based (IRT-based) indices (OUTFITz, INFITz, ECI2z, ECI4z, and lz). Four kinds of comparative conditions, including content category (CC), types of aberrance (AT), severity of aberrance (AS), and the ratios of aberrant persons (AP), were implemented under the tolerance of a .05 false positive rate. Results showed that group-based indices performed better than IRT-based indices. Although the BW indices and most of the other group-based indices exhibited over 90% detection rates, the BW indices exhibited the best stability across implemented conditions. On the basis of their highly stable detection power and objective cutoffs, the BW person-fit indices were recommended for use in diagnosing students’ learning issues in classrooms.

Keywords
Person-fit index, BW indices, group-based indices, IRT-based indices, detection power

Introduction

The indices that have been developed to detect aberrant response patterns are referred to as unusual response indicators, caution indices, fit indices, aberrance indices, appropriateness measurement indices, or likelihood indices (Meijer & Sijtsma, 1995; D’Costa, 1993a, 1993b). According to measurement theory, some indices are group based, and some are based on item response theory (IRT; Harnisch & Linn, 1981; Kogut, 1986, Meijer & Sijtsma, 1999). Group-based indices refer to those indices that use certain group characteristics (e.g., the concept of item difficulty or the proportion of correct item responses to the total number of responses) to identify aberrances. On the other hand, most IRT-based indices measure the degree of consistency of an observed response pattern with respect to a certain IRT model used.

Most group-based aberrance indices, however, encounter the problem of without knowing their theoretical distributions such that some alternative approaches, like rules of thumb, are provided to enable clinical use. For example, the original Sato caution index (SCI; Sato, 1975) was deemed as aberrant when it was higher than 0.5. Harnisch and Linn (1981) later proposed the modified caution index (MCI) values and suggested that values greater than 0.3 should be considered aberrant. For the within-ability-concern and beyond-ability-surprise indices (Wc&Bs) which were introduced by D’Costa (1993a, 1993b), values between 0.3 and 0.5 necessitated “routine caution,” and values greater than 0.5 required “serious caution.” Consequently, the lack of absolute cut-off standards results in the sample-dependent identification and interpretation of aberrant responses, and the term of “index” may even be questioned for these indices. On the other hand, although some IRT-based indices have been standardized in order to examine their null-hypothesis-based distributions, the challenge of approximating corresponding distributions—usually normal distributions—still exists under the assumption of large samples. In other words, the use of these indices is only appropriate for large samples and cannot be guaranteed to be appropriate for small samples. Therefore, it is questionable to use these asymptotic-distribution-based indices to infer aberrance when sample sizes are small, especially when the data sets do not have normal distributions.

Two group-based indices, the beyond-ability-surprise index (B) and the within-ability-concern index (W) both inheriting from the Wc&Bs indices, can apply cut-off standards in small samples (Huang, 2007). The B index was designed to detect the “beyond-ability” aberrant response patterns and the W index detected those that are “within-ability.” The beyond-ability response pattern is assessed using a Guttman scale: it measures the surprise of a person when he/she correctly answers items beyond his/her ability level. Someone exhibiting a within-ability response pattern is considered to need more attention because some of their wrong answers are below their ability levels. The aberrances cutoffs provided by the BW indices are based on a permutation technique that is norm-referred for each ability-ratio/error-ratio (or T/K-E/K) cell (Huang, 2007). They are established according to various ability ratios and error ratios under three types of percentiles (90%, 95%, and 99%). Any observed B or W values greater than the
cutoff in a particular cell are judged “aberrant” under a particular percentile comparison. It is essential to mention that each cell contains cutoffs based on simulated occurrences of index values. Thus, a statistically significant aberrance under a specific level of positive false rate is based on the comparison with other persons’ aberrant performances. However, the relative powers of the BW indices have not been disclosed yet.

Typically, the power of indices is based on the rate of detecting aberrances. Meijer and Sijtsma (1995, 1999) had argued that the power of detecting aberrances by IRT-based indices is better than that of group-based indices. However, some researchers continued to highly recommend group-based indices. For example, Karabatsos (2003) examined thirty-six developed indices, including IRT-based and group-based indices. He found that group-based indices detected aberrances more effectively than IRT-based indices did. Thus, it is very valuable to compare the aberrance-detecting power of the BW indices with that of other group-based and IRT-based indices.

On the basis of the above reasons, this study compares the detection power of the BW indices with nine other famous aberrance indices under different conditions. Four of them are group-based indices, including SCI (Sato, 1975), MCI (Harnisch & Linn, 1981), norm conformity index (NCI; Tatsuoka & Tatsuoka, 1982), and Wc&Bs (D’Costa, 1993a, 1993b). The remaining five are IRT-based indices, including OUTFITz and INFITz (Smith, 1991, Linacre & Wright, 1994), ECI2z and ECI4z (Tatsuoka & Linn, 1983), and Iz (Drasgow, Levine, & Williams, 1985).

BW indices

The BW indices are identified by comparing individuals’ score patterns with the perfect Guttman pattern, which requires individuals to have answered a difficult item correctly, then they must respond correctly those items with difficulty levels lower than the difficult item, given the order of item difficulty is arranged from easy to hard (Guttman, 1944). On the basis of the “discrepancy distance” concept, the distance between one’s corrected ability and the difficulty of an item with an unexpected response can be calculated. The corrected ability for a T score person is defined as the mean of the difficulties of the \(i^{th}\) and the \((i+1)^{th}\) items. In other words, corrected ability references the total raw score to a continuous ITEM difficulty scale on which an examinee’s ability can be taken into account. Then, the sums of the discrepancy distances calculated within and beyond the examinee’s ability level serve as the numerators of the W and B indices, respectively. These two kinds of discrepancy distances are adjusted and bounded by the denominator with the concept of “maximum discrepancy distance.” The person-fit BW indices (Huang, 2006, 2007) are defined as follows:

\[
W_i = \sum_{j=1}^{K} (1-u_{ij}) \times \left( \frac{q_{ij}^* - q_j}{(K-1)/2} \right) \times 100
\]

\[
B_i = \sum_{j=1}^{K} u_{ij} \times \left( \frac{q_{ij}^* - q_j}{(K-1)/2} \right) \times 100
\]

where \(u_{ij}\) represents responses: 1 for correct answers and 0 for incorrect answers. The \(q\) variables represent the levels of item difficulty ordered from easy to hard and are bounded within the interval of [0, 1]; \(q_{ij}^*\) is the corrected ability level for the T score person. The bracketed expression with the test length \(K\), \([(K - 1)/2]\), is the theoretical maximum value of the numerator and is equal to the lower Gauss integer and expressed by the smallest integer greater than or equal to the value of \((K - 1)/2\). Finally, the multiplication of 100 to the entire expression of the fraction is for the convenience of providing easily understandable values for the two indices. The values of the two indices will be greater than 0.

Some characteristics of the BW indices have been revealed recently. They include the norm-referred cutoffs to identify aberrances under different data matrices (Huang, 2007, 2008), robustness against test length (Huang, 2010), and diagnostic spaces to classify those students who have similar misconception types (Huang, in press). In addition, the BW indices can differentiate varying degrees of aberrant response patterns in the same test scores in a small sample size class (Huang, 2006). The indices were also applied to empirically predict number sense performances for sixth grade elementary school students (Yeh, Yang, & Huang, 2006) as well as to the Basic Competence Test for junior high school students in Taiwan (Tsai, 2010).
Detection power

Typically, the relative power of detecting aberrances is a well-known criterion for evaluating aberrance indices. The higher the detection rate of aberrance indices at a given false positive rate, the more powerful is the index. Some studies have compared several aberrance indices using simulation data or real data. Harnisch and Linn (1981) compared 10 aberrance indices using empirical assessment data from the Illinois Inventory of Educational Progress (IIEP). They found that although MCI displayed the least correlation with total score, the SCI displayed a negative correlation with total score. Drasgow, Levine, and McLaughlin (1987) examined standardized forms of nine indices across three levels of ability and found that the SCI was not stable between the low ability level and high/average ability levels. D’Costa (1993b) used simulation data to examine the distributions of Wc&Bs and MCI for different values of total score. He found that the MCI performed well in identifying aberrant responses when the number of spuriously high/low total scores increased; his findings were similar to Runder’s findings (1983). Recently, Lu, Huang, and Fan (2007) compared the four Guttman-based indices, (i.e., the SCI, MCI, NCI, and Wc&Bs indices). They found that detection rates were significantly different among the four indices under their original thumb-rule cutoffs (in descending order, Wc&Bs, SCI, MCI, and NCI); however, under modified cutoffs at the 80th percentile of permutations, the order was slightly different (in descending order, Wc&Bs, SCI, NCI, and MCI).

In the comparisons of IRT-based indices, Birenbaum (1985) found that ECI2z, ECI4z, and lz were superior with respect to their low correlations to the total score and their high capability of detecting aberrant response patterns. Drasgow et al. (1987) examined nine IRT-based indices and found that lz and ECI2z provided higher detection rates than ECI4z in some forms of aberrances, but the results were not consistent in other forms. Noonan, Boss, and Gessaroli (1992) compared lz, ECI4z, and non-standardized INFIT indices under the conditions of an IRT model, test length, and three false positive rates. They found ECI4z showed the best performance and least affected by test length and the IRT model, whereas non-standardized INFIT index was most affected by the implemented conditions. Li and Olejnik (1997) examined five Rasch aberrance indices (ECI2z, ECI4z, lz, OUTFITz, and INFITz). They found ECI2z, ECI4z, lz, INFITz performed equally well in detecting aberrances regardless of test dimensionality, type of misfit, and test length. They also found the five Rasch-based indices were more sensitive to spuriously high responses than to spuriously low ones in a two-dimensional test, and the detectability of aberrance increased with test length. Similar research was performed in Seol’s study (1998). She compared these five standardized Rasch-model-based aberrance indices and found the ECI2z and ECI4z indices appeared more sensitive to the presence of guessing and carelessness than the other three indices.

From previous studies, it seems that group-based indices (SCI, MCI, Wc&Bs, NCI) and IRT-based indices (ECI2z, ECI4z, lz, OUTFITz, INFITz) can be good alternatives to the BW indices. In addition, it was found that the power of an index was usually specified by its detection rate under different conditions. Various conditions were commonly used, such as content category (Drasgow, Levine, & McLaughlin, 1991; Harnisch & Linn, 1981; Li & Olejnik, 1997; Meijer, 1997; Reise, 1995; Schmitt, Cortina, & Whitney, 1993), types of aberrances (Drasgow et al., 1991; Meijer, Molenaar, & Sijtsma, 1994; Meijer, 1997; Nering & Meijer, 1998), and severity of aberrances (Drasgow & Levine, 1986; Drasgow et al., 1991; Meijer, 1997; Nering & Meijer, 1998; Rudner, 1983). “Content category” refers to the content of a test. “Types of aberrances” refers to either spuriously high, spuriously low score aberrances or both simultaneously. “Severity of aberrance” refers to the percentage of spurious responses implemented in a normal response pattern. In addition, the condition of the ratio of aberrant examinees would be considered in this study because it is rare in the literature and because the aberrant responses may become non-aberrant when we set the severity of aberrance for all examinees. Thus, four kinds of comparative conditions, content category (CC), types of aberrance (AT), severity of aberrance (AS), and the ratio of aberrant person (AP), would be implemented under the tolerance of .05 false positive rates in the current study.

Method

Design

Empirical data from Huang’s test (2003) of elementary school students’ mathematical representative ability including the graph representation subscale (GR) and the symbol representation subscale (SR), were used in this study. Both subscales contained 16 items. There were an average of 33 students in each class, and 498 students were dispersed among 15 classes. The data matrices of students’ responses in each class served as simulated seeds. However, if responses were ranked by item difficulty and person ability after permuting data by columns, most of
responses of “1” would fall to the bottom-right corner of the matrix. This would lead to a significant difference between the simulated matrix and the original data matrix. In order to resolve this problem, the weighted-permutation indicator (Lu, Huang, & Fan, 2007) was used by multiplying each examinee’s total score to its fourth power with a random probability. The results showed that no difference exists between a simulated matrix and its corresponding original data matrix.

Aside from the content categories, three other conditions were controlled in this study: AT, AS, and AP. In the AT condition, the following three types of aberrances were designed in response patterns: the spuriously high score response pattern (AT_h), the spuriously low score response pattern (AT_l), and both spuriously high and low score response pattern (AT_b). Three levels of aberrant severity (10%, 20%, and 30%) were also designed in the AS condition. Regarding the AP condition, the levels of AP determined how many people were selected to serve as spuriously high score and spuriously low score aberrant examinees. Three ratio levels of aberrant persons (10%, 20%, and 30%) were set in the three spurious aberrances.

In order to ensure that no full or null scores were generated, randomly selected examinees were chosen from those persons whose correct responses were within 20% and 80% of the total items. For the AT_h patterns, examinees were randomly selected from those who had correctly answered 80%, 70%, and 60% of the total items, and 10%, 20%, and 30% of their responses were directly changed to “1” from right to left, irrespective of what their original responses were. Similarly, the AT_l patterns were generated by changing 10%, 20%, and 30% of the responses from left to right to “0” for those who had correctly answered 20%, 30%, and 40% of the total items, irrespective of what their original responses were. Finally, those persons who had correctly answered 10%-90%, 15%-85%, and 20%-80% of the total items were randomly selected as examinees, and the AT_b patterns were generated by changing both the right-to-left and left-to-right 5%, 10%, and 15% of responses to 1 and 0, respectively, irrespective of what their original responses were.

Process and Analysis

All group-based indices were calculated using the WBStar program (Lu & Huang, 2010). All IRT-based indices based on the two-parameter IRT model (for the Rasch model, the discrimination was set as “1”) were estimated by the BILOG-MG program (Zimowski, Muraki, Mislevy, & Bock, 1996), where the parameters of items were estimated by the marginal maximum likelihood estimation (MML) method, and the person ability parameter was estimated by the expected a posteriori estimation (EAP) method. Ten steps were executed by the Aberrance Indices Program for Simulation (AIPS; written using Visual Basic; Lu & Huang, 2006): (1) read an empirical data matrix from each of the fifteen classes, including the graph and symbol representation subscales; (2) set manipulating conditions (AT, AS, and AP) in the original empirical data matrix; (3) use a weighted-permutation technique (mentioned above) to generate a simulated data matrix corresponding to its original data matrix; (4) change response codes according to various conditions of AT, AS, and AP in the original empirical data matrix; (5) use a weighted-permutation technique (mentioned above) to generate a simulated data matrix corresponding to its original data matrix; (6) calculate detection rates for each aberrant index according to the cutoffs of indices; (7) replicate Steps 3–6 to generate 100 permutations; (8) repeat Steps 2–7 under 27 manipulated conditions (3ATs × 3As × 3APs); (9) repeat Steps 1–8 for fifteen classes to generate 81,000 simulated aberrant responses (2 content categories × 27 conditions × 100 permutations × 15 classes); and finally, (10) analyze the outputs of detection rates for indices.

Due to the characteristics of separation inherited from the Wc&Bs indices, the estimations of the BW indices were not like other group- or IRT-based indices, which use a person’s entire responses to estimate index values. Instead, the BW indices were estimated on the basis of a part of an individual’s responses, that is, within or beyond a person’s ability. Thus, it was appropriate to calculate the units of aberrance for these two indices by counting those whose response patterns reached the norm-referred cutoffs of within-ability aberrances or beyond-ability aberrances (Huang, 2007). Concerning the other group-based indices, the criteria of aberrances were greater than .5 for SCI and Wc&Bs, greater than .3 for MCI, and less than 0 for NCI. The criteria of aberrances for all standardized IRT-based indices were set as greater than 1.96.

Therefore, by putting all the detection rates in all the conditions, the question “What are the differences in detection rates among these ten aberrance indices?” can be answered by a five-way (including the index itself) ANOVA statistic approach. The question “Are these indices stable across each manipulated condition?” can be answered by a four-way MANOVA statistic approach with the Tamhane’s T2 post hoc comparisons (a conservative pairwise-comparison test based on a t test as the homogeneity assumption was rejected).
Results

Detection rates

As seen in Table 1, although all factorial interactions were significant at the .001 alpha level, most of the interactive effect sizes were small (partial $\eta^2$’s < .10). The maximal effect size occurred in the main effects of the index (partial $\eta^2 = .69$). This implies the index factor contributed to the difference in detection rates the most. Therefore, we examined the difference in detection rates among indices. In Figure 1, the group-based indices seemed to perform better than the IRT-based indices. The detection rates of all group-based indices, except for the NCI, were over .90. The detection rates of all the IRT-based indices, except for the ECI2z and the INFITz were over .50 but lower than .70. Note that with a detection rate of .92, the BW performed as well as Wc&Bs, SCI, and MCI did.

According to the data in Table 1, the condition factors showed interactive effects on detection rates, but their effects were very small (all were less than .02, except for the effects of the index factor on the AT and AS factors, both partial $\eta^2 = .10$). Thus, it was appropriate to investigate the stability of indices in the AT and AS conditions individually and disregard the other interactive effects between these factors.

<table>
<thead>
<tr>
<th>Sources</th>
<th>Type III SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index (IN)</td>
<td>97517.79</td>
<td>9</td>
<td>10835.31</td>
<td>195934.65***</td>
<td>.69</td>
</tr>
<tr>
<td>Content Category (CC)</td>
<td>22.53</td>
<td>1</td>
<td>22.53</td>
<td>407.48***</td>
<td>.00</td>
</tr>
<tr>
<td>Aberrance Type (AT)</td>
<td>3949.63</td>
<td>2</td>
<td>1974.81</td>
<td>35710.52***</td>
<td>.08</td>
</tr>
<tr>
<td>Aberrance Severity (AS)</td>
<td>3859.39</td>
<td>2</td>
<td>1929.70</td>
<td>34894.67***</td>
<td>.08</td>
</tr>
<tr>
<td>Aberrance Person (AP)</td>
<td>418.88</td>
<td>2</td>
<td>209.44</td>
<td>3787.27***</td>
<td>.01</td>
</tr>
<tr>
<td>IN × CC</td>
<td>685.96</td>
<td>9</td>
<td>76.22</td>
<td>1378.25***</td>
<td>.02</td>
</tr>
<tr>
<td>IN × AT</td>
<td>4928.06</td>
<td>18</td>
<td>273.78</td>
<td>4950.78***</td>
<td>.10</td>
</tr>
<tr>
<td>IN × AS</td>
<td>4822.23</td>
<td>18</td>
<td>267.90</td>
<td>4844.46***</td>
<td>.10</td>
</tr>
<tr>
<td>IN × AP</td>
<td>393.93</td>
<td>18</td>
<td>21.89</td>
<td>395.75***</td>
<td>.01</td>
</tr>
<tr>
<td>CC × AT</td>
<td>278.80</td>
<td>2</td>
<td>139.40</td>
<td>2520.80***</td>
<td>.01</td>
</tr>
<tr>
<td>CC × AS</td>
<td>1002.87</td>
<td>2</td>
<td>501.43</td>
<td>9067.42***</td>
<td>.02</td>
</tr>
<tr>
<td>CC × AP</td>
<td>15.09</td>
<td>2</td>
<td>7.54</td>
<td>136.42***</td>
<td>.00</td>
</tr>
<tr>
<td>AT × AS</td>
<td>54.64</td>
<td>4</td>
<td>13.66</td>
<td>247.02***</td>
<td>.00</td>
</tr>
<tr>
<td>AT × AP</td>
<td>15.89</td>
<td>4</td>
<td>3.97</td>
<td>71.85***</td>
<td>.00</td>
</tr>
<tr>
<td>AS × AP</td>
<td>24.91</td>
<td>4</td>
<td>6.23</td>
<td>112.60***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × CC × AT</td>
<td>550.84</td>
<td>18</td>
<td>30.60</td>
<td>553.38***</td>
<td>.01</td>
</tr>
<tr>
<td>IN × CC × AS</td>
<td>614.83</td>
<td>18</td>
<td>34.16</td>
<td>617.66***</td>
<td>.01</td>
</tr>
<tr>
<td>IN × CC × AP</td>
<td>119.45</td>
<td>18</td>
<td>6.64</td>
<td>120.00***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × AT × AS</td>
<td>305.26</td>
<td>36</td>
<td>8.48</td>
<td>153.33***</td>
<td>.01</td>
</tr>
<tr>
<td>IN × AT × AP</td>
<td>43.17</td>
<td>36</td>
<td>1.20</td>
<td>21.68***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × AS × AP</td>
<td>125.10</td>
<td>36</td>
<td>3.47</td>
<td>62.84***</td>
<td>.00</td>
</tr>
<tr>
<td>CC × AT × AS</td>
<td>34.02</td>
<td>4</td>
<td>8.51</td>
<td>153.80***</td>
<td>.00</td>
</tr>
<tr>
<td>CC × AT × AP</td>
<td>6.37</td>
<td>4</td>
<td>1.59</td>
<td>28.80***</td>
<td>.00</td>
</tr>
<tr>
<td>CC × AS × AP</td>
<td>23.01</td>
<td>4</td>
<td>5.75</td>
<td>104.02***</td>
<td>.00</td>
</tr>
<tr>
<td>AT × AS × AP</td>
<td>20.70</td>
<td>8</td>
<td>2.59</td>
<td>46.80***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × CC × AT × AS</td>
<td>139.21</td>
<td>36</td>
<td>3.87</td>
<td>69.93***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × CC × AT × AP</td>
<td>43.48</td>
<td>36</td>
<td>1.21</td>
<td>21.84***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × CC × AS × AP</td>
<td>39.31</td>
<td>36</td>
<td>1.09</td>
<td>19.74***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × AS × AT × AP</td>
<td>53.39</td>
<td>72</td>
<td>0.74</td>
<td>13.41***</td>
<td>.00</td>
</tr>
<tr>
<td>CC × AS × AT × AP</td>
<td>3.50</td>
<td>8</td>
<td>0.44</td>
<td>7.92***</td>
<td>.00</td>
</tr>
<tr>
<td>IN × CC × AS × AT × AP</td>
<td>22.32</td>
<td>72</td>
<td>0.31</td>
<td>5.61***</td>
<td>.00</td>
</tr>
<tr>
<td>Error</td>
<td>44763.65</td>
<td>809460</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>467477.18</td>
<td>810000</td>
<td></td>
<td></td>
<td>.00</td>
</tr>
</tbody>
</table>

*p < .001
Stability of Indices

Table 2 shows a summary of one-way MANOVA results of indices for the AT condition (Wilk’s Lambda = .690, p < .001, partial $\eta^2 = .169$). As evident from the table, except for the Wc&Bs indices, all indices performed more sensitively on the “both spuriously high and low” aberrances than on the “spuriously high” or “spuriously low” aberrances. Note that three IRT-based indices, $l_z$, $ECI_4z$, and $OUTFITz$, varied greatly under this condition with partial $\eta^2$ values of .21, .13, and .23, respectively. This implies that these three IRT-based indices are more unstable across different types of aberrances than the five group-based indices did. Although $ECI_2z$ and $INFITz$ exhibited the lowest effect size, their detection rates were still very small. This does not make much sense. With respect to the group-based indices, the BW indices exhibited similar detection rates to Wc&Bs for spuriously “high,” “low,” and “both” score aberrances ($Mean = .91, .89, and .95$ respectively). Since the BW indices exhibited a relatively low effect size (partial $\eta^2 = .01$), the detection rate differences among aberrant types did not have much practical meaning. This indicated that the BW indices were stable across the three types of aberrances.

Table 2. One-way MANOVA for the types of aberrances (AT) condition (N = 81,000)

<table>
<thead>
<tr>
<th>Index</th>
<th>AT</th>
<th>Mean</th>
<th>SD</th>
<th>df</th>
<th>F</th>
<th>partial $\eta^2$</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCI</td>
<td></td>
<td>.97</td>
<td>.12</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>SCI</td>
<td>2</td>
<td>.98</td>
<td>.10</td>
<td>2</td>
<td>1288.12***</td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>SCI</td>
<td>3</td>
<td>.92</td>
<td>.21</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>MCI</td>
<td>1</td>
<td>.96</td>
<td>.15</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>MCI</td>
<td>2</td>
<td>.96</td>
<td>.16</td>
<td>2</td>
<td>2993.40***</td>
<td>.07</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>MCI</td>
<td>3</td>
<td>.84</td>
<td>.30</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>NCI</td>
<td>1</td>
<td>.64</td>
<td>.40</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>NCI</td>
<td>2</td>
<td>.71</td>
<td>.39</td>
<td>2</td>
<td>755.71***</td>
<td>.02</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>NCI</td>
<td>3</td>
<td>.57</td>
<td>.42</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>Wc&amp;Bs</td>
<td>1</td>
<td>.99</td>
<td>.05</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>Wc&amp;Bs</td>
<td>2</td>
<td>.98</td>
<td>.11</td>
<td>2</td>
<td>258.35***</td>
<td>.01</td>
<td>High &gt;Both= Low</td>
</tr>
<tr>
<td>Wc&amp;Bs</td>
<td>3</td>
<td>.98</td>
<td>.11</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>Wc&amp;Bs</td>
<td>1</td>
<td>.91</td>
<td>.23</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>BW</td>
<td>2</td>
<td>.95</td>
<td>.17</td>
<td>2</td>
<td>519.23***</td>
<td>.01</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>BW</td>
<td>3</td>
<td>.89</td>
<td>.24</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>BW</td>
<td>1</td>
<td>.77</td>
<td>.35</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>$l_z$</td>
<td>2</td>
<td>.79</td>
<td>.35</td>
<td>2</td>
<td>10629.56***</td>
<td>.21</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>$l_z$</td>
<td>3</td>
<td>.39</td>
<td>.40</td>
<td></td>
<td></td>
<td>.03</td>
<td>Both&gt;High&gt;Low</td>
</tr>
<tr>
<td>$ECI_2z$</td>
<td>1</td>
<td>.00</td>
<td>.04</td>
<td>2</td>
<td>330.30***</td>
<td>.01</td>
<td>Both&gt;High&gt;Low</td>
</tr>
</tbody>
</table>
Note: a “spuriously high” aberrances, b “both spuriously high and low” aberrances, and c “spuriously low” aberrances. ***p < .001.

Regarding the condition of aberrant severity, Table 3 shows a summary of the results of a one-way MANOVA for the indices under the AS condition (Wilk’s Lambda = .604, p < .001, $\eta^2 = .223$). The NCI and ECI4z indices had large effect sizes (partial $\eta^2 = .34$, and .12, respectively). This may indicate that these two indices were more influenced by different levels of aberrant severity than other indices. Even though ECI2z and INFITz exhibited the lowest effect sizes of all indices, their detection rates were very low. This does not make much sense. On the other hand, the BW indices exhibited high detection rates (.95, .92, and .89 for different levels of aberrance severity, respectively) but low effect size (partial $\eta^2 = .01$). This indicates that the differences in detection rates among levels of aberrant sensitivity did not contain much practical meaning; furthermore, it reveals the strong detection power and stability of the BW indices across different levels of aberrant severity.

### Table 3. One-way MANOVA for the aberrant severity (AS) condition (N = 81,000)

<table>
<thead>
<tr>
<th>Index</th>
<th>AS</th>
<th>Mean</th>
<th>SD</th>
<th>df</th>
<th>F</th>
<th>partial $\eta^2$</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCI</td>
<td>1a</td>
<td>.90</td>
<td>.23</td>
<td>2</td>
<td>2822.87***</td>
<td>.07</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>.98</td>
<td>.10</td>
<td>2</td>
<td>3271.02***</td>
<td>.07</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3c</td>
<td>.99</td>
<td>.06</td>
<td>2</td>
<td>20917.06***</td>
<td>.34</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.83</td>
<td>.30</td>
<td>2</td>
<td>1990.97***</td>
<td>.05</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>MCI</td>
<td>2</td>
<td>.95</td>
<td>.17</td>
<td>2</td>
<td>515.97***</td>
<td>.01</td>
<td>10% &gt;20% &gt;30%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.97</td>
<td>.13</td>
<td>2</td>
<td>3416.95***</td>
<td>.08</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.32</td>
<td>.37</td>
<td>2</td>
<td>1987.38***</td>
<td>.05</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>NCI</td>
<td>2</td>
<td>.70</td>
<td>.36</td>
<td>2</td>
<td>3416.95***</td>
<td>.08</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.89</td>
<td>.24</td>
<td>2</td>
<td>5782.52***</td>
<td>.12</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.95</td>
<td>.16</td>
<td>2</td>
<td>515.97***</td>
<td>.01</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>BW</td>
<td>2</td>
<td>.92</td>
<td>.21</td>
<td>2</td>
<td>1990.97***</td>
<td>.05</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.89</td>
<td>.26</td>
<td>2</td>
<td>515.97***</td>
<td>.01</td>
<td>10% &gt;20% &gt;30%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.49</td>
<td>.42</td>
<td>2</td>
<td>3416.95***</td>
<td>.08</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>$l_z$</td>
<td>2</td>
<td>.70</td>
<td>.39</td>
<td>2</td>
<td>3416.95***</td>
<td>.08</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.76</td>
<td>.37</td>
<td>2</td>
<td>3416.95***</td>
<td>.08</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.01</td>
<td>.07</td>
<td>2</td>
<td>3.85</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>ECI2z</td>
<td>2</td>
<td>.01</td>
<td>.06</td>
<td>2</td>
<td>3.85</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.01</td>
<td>.07</td>
<td>2</td>
<td>3.85</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.30</td>
<td>.37</td>
<td>2</td>
<td>3.85</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>ECI4z</td>
<td>2</td>
<td>.05</td>
<td>.41</td>
<td>2</td>
<td>5782.52***</td>
<td>.12</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.65</td>
<td>.41</td>
<td>2</td>
<td>5782.52***</td>
<td>.12</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>2</td>
<td>22.84***</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>INFITz</td>
<td>2</td>
<td>.00</td>
<td>.00</td>
<td>2</td>
<td>22.84***</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.00</td>
<td>.00</td>
<td>2</td>
<td>22.84***</td>
<td>.00</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
<tr>
<td>OUTFITz</td>
<td>1</td>
<td>.41</td>
<td>.41</td>
<td>2</td>
<td>1987.38***</td>
<td>.05</td>
<td>30% &gt;20% &gt;10%</td>
</tr>
</tbody>
</table>
Note: *10% aberrant severity, b20% aberrant severity, and c30% aberrant severity.***

### Conclusion

In conclusion, group-based indices seemed to perform better than IRT-based indices. The Wc&Bs, SCI, MCI, and BW indices seemed to dominate the other indices across all conditions. The NCI, Iz, ECI4z, and OUTFITz indices exhibited mediocre performance, and the ECI2z and INFITz indices had the lowest detection rates.

Generally, the findings of the superiority of the group-based indices over the IRT-based indices are consistent with the study comparing thirty-six indices by Karabatsos (2003). The superior detection power of Wc&Bs is supported by the study of Lu, Huang, and Fan (2007). The good performance of the MCI is consistent with the findings of the studies by D’Costa (1993b) and Rudner (1983). Regarding the IRT-based indices, the finding of good performance of Iz is consistent with the findings of Birenbaum (1985), Drasgow et al. (1987), and Li and Olejnik (1997). Although the consistently good performances of ECI4z is consistent with the findings of Birenbaum (1985), Noonan, Boss, and Gessaroli (1992), Li and Olejnik (1997), and Soel (1998), it is not supported by Drasgow et al. (1987). The poor performance of ECI2z in this study is not consistent with other literatures, for example, Birenbaum (1985), Drasgow et al. (1987), Li and Olejnik (1997), and Soel (1998). With respect to the examinations of indices stability in the study, Iz, ECI4z, and OUTFITz seemed unstable across the three aberrance type conditions. The NCI and ECI4z indices were unstable across the three severity level conditions. This is consistent with previous literatures, for example, Lu, Huang, and Fan (2007), Drasgow et al. (1987), Li and Olejnik (1997), and Soel (1998). This seems to indicate that these indices were condition-based. However, the BW indices exhibited the most stability across the AT and AS conditions.

The reason the group-based indices performed better than the IRT-based indices may be that group-based indices are more response pattern oriented rather than response probability oriented like the IRT-based indices. When we permuted data from the original matrix, we changed people’s response patterns, but this may not have changed the probability of someone answering an item correctly. The parameters estimated in the group-based indices are based on other people’s relative response patterns, and therefore would be sensitive to changes in response patterns. On the other hand, the parameters estimated in the IRT-based indices are absolute across items and across persons; they are not as sensitive to changes in response patterns. Thus, when aberrant conditions were implemented, the IRT-based indices were less effective than group-based indices.

Specifically, the reason for the poor performances of ECI2z may be due to its formula device. ECI2 measures the similarity of an observed pattern to group probabilities of correct answers. In this study, there were only approximately 33 examinees in a class responding to 16 items, and the person-fit indices were estimated on the basis of the unit data matrix at one time. The small sample size may have resulted in insignificant changes in the central ordered responses. Individual response patterns may have been similar to the group response patterns. Thus, the covariance of the observed response vector for a person and the vector for group probabilities for correct answers may be large in this study. This would lead to small values of ECI2z and result in its insensitivity when detecting aberrances. In contrast, ECI4z measures the similarity of an observed pattern to individual probabilities of correct answers. Individual probabilities for correct answers were measured through an IRT model; thus, ECI4z was less influenced by ordered small sample sizes.

On the other hand, it was also interesting to note the contrast between OUTFITz and INFITz. Linacre and Wright (1994) provide a possible reason: OUTFITz is outlier-sensitive and dominated by unexpected outliers. INFITz is dominated by unexpected inlying patterns and is inlier-sensitive. Due to the small sample size and short item situation, a lack of significant changes in the central ordered responses might not lead to many unexpected inlying patterns; thus, INFITz was not sensitive. As expected, OUTFITz is outlier-sensitive, which fit well with this study because unexpected outliers occurred frequently.

On the basis of the above findings, group-based indices, at least those of Wc&Bs, SCI, MCI, and BW in this study, must not be overlooked. They outperformed famous IRT-based indices due to their superior detection powers and
their easily understandable devices. No complicated calculations were needed in their estimations, and they were always sensitive to changes in response patterns. In other words, they provide a more accurate reflection of the changes in people’s response patterns. Moreover, unlike the IRT-based indices, they are suitable for use in small samples, such as students in one class. However, the cutoffs settings of group-based indices still necessitate caution. As mentioned in the problem statement, the cutoffs of group-based indices (except for BW indices) are based on a certain empirical data or rules of thumb. Subjective criteria for cutoffs would cause the thresholds for detecting aberrances to be reached too easily, and “spurious” high detection rates may occur. Instead, the BW indices performed as well as Wc&Bs, SCI, and MCI; they had good detection rates and outperformed other IRT-based indices. They also exhibited the most stability across the AT and AS conditions among all indices. In addition, due to their sensitivity toward changes in people’s response patterns (Huang, 2006) and their established objective cutoffs (Huang, 2007), the BW can provide more conservative and reliable results for small sample sizes. The BW indices are strongly recommended for teachers who wish to diagnose students’ learning in class. Teachers may realize who tends to guess or tends to slip through the B index and the W index, respectively. A student exhibits a high value of the B index indicates he or she may obtain a “spuriously high” score that may be attributed by guessing or by creative thinking; in contrast, a student with high values of the W index may need more concerns about his/her carelessness. This study still has a few limitations. First, the AP factor manipulated in this study showed slight effect on the indices. The reason might be due to the selection of those persons whose correct responses were within 20% and 80% of the total items in order to ensure no full or null scores were generated. One might use the entire persons in the future. Second, only the AT, AS, and AP conditions were manipulated in this study. In order to ensure other sources of an index’s detection power, one can add other factors, such as item length and sample size, and other IRT models, in future studies. Finally, detection power comparisons among aberrance indices in this study are based on “spuriously aberrance response patterns.” In order to authentically reflect examinees’ response patterns, one might compare the detection power based on “true response patterns.”

Acknowledgements

The author would like to thank the National Science Council (NSC) in Taiwan for financial support and sincere appreciations would be expressed to Journal Reviewers for their helpful recommendations.

References


Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios

Dirk Ifenthaler
Fakultät für Sozialwissenschaften, Universität Mannheim, D-68159 Mannheim, Germany // ifenthaler@uni-mannheim.de

ABSTRACT
Cognitive scientists have studied internal cognitive structures, processes, and systems for decades in order to understand how they function in human learning. In order to solve challenging tasks in problem situations, learners not only have to perform cognitive activities, e.g., activating existing cognitive structures or organizing new information, they also have to set specific goals, plan their activities, monitor their performance during the problem-solving process, and evaluate the efficiency of their actions. This paper reports an experimental study with 98 participants where effective instructional interventions for self-regulated learning within problem-solving processes are investigated. Furthermore, an automated assessment and analysis methodology for determining the quality of learning outcomes is introduced. The results indicate that generic prompts are an important aid for developing cognitive structures while solving problems.

Keywords
Reflection, metacognition, prompting, HIMATT

Introduction
Self-regulated learning is regarded as one of the most important skills needed for life-long learning. Zimmerman (1989, p. 4) describes self-regulated learning as a process in which learners “are metacognitively, motivationally, and behaviorally active participants in their own learning process.” Hence, self-regulated learning is a complex process which involves numerous dimensions of human information processing (Azevedo, 2008, 2009; Pintrich, 2000; Schraw, 2007; Veenman, van Hout-Wolters, & Afflerbach, 2006; Zimmerman, 2008). Accordingly, in order to solve challenging tasks in problem situations, learners not only have to perform cognitive activities, e.g., activating existing knowledge structures or organizing new information (Seel, Ifenthaler, & Pirnay-Dummer, 2009), they also have to set specific goals, plan their activities, monitor their performance during the problem-solving process, and evaluate the efficiency of their actions (Wirth & Leutner, 2008).

Moreover, the facilitation of self-regulated learning is a balancing act between necessary external support and desired internal regulation (Koedinger & Aleven, 2007; Simons, 1992). From an instructional point of view, there are two vital ways to externally support self-regulated learning within problem-solving processes. Direct external support, in terms of direct instruction, aims at facilitating explicit problem-solving strategies and skills as well as their application and transfer to different domains. Hence, direct instruction could include detailed scaffolds (step-by-step instruction) on how to solve a specific phenomenon in question (Collins, Brown, & Newman, 1989). Indirect external support provides learning aids which induce and facilitate already existing problem-solving strategies and skills. Accordingly, if learners already possess comprehensive problem-solving strategies but fail to use this knowledge in a specific situation, it seems reasonable to motivate them to apply their existing strategic knowledge effectively (Lin & Lehmann, 1999). A possible instructional method for indirectly guiding and supporting the regulation of learners’ problem-solving processes is prompting (Wirth, 2009). In general, prompts are presented as simple questions (e.g., “What will be your first step when solving the problem?”), incomplete sentences (e.g., “To approach the solution to the problem step by step, I have to …”), explicit execution instructions (e.g., “First, draw the most important concepts and link them.”), or pictures and graphics for a specific learning situation (Bannert, 2009). Accordingly, well-designed and embedded prompts direct learners to perform a specific desired activity which is contextualized within a particular problem-solving situation (see Davis, 2003; Davis & Linn, 2000; Lin & Lehmann, 1999). According to Davis (2003), prompts can be categorized into generic and directed prompts. While the generic prompt only asks learners to stop and reflect about their current problem-solving activities, the directed prompt also provides them with an expert model of reflective thinking in the problem-solving process.

From a methodological point of view, we argue that it is essential to identify economic, fast, reliable, and valid techniques to assess and analyze these complex problem-solving processes. Especially in experimental settings where huge sets of data need to be processed, standard methodologies (e.g., paper and pencil tests) may have
disadvantages with regard to analysis economy. Therefore, we developed an automated assessment and analysis technology, HIMATT (Highly Integrated Model Assessment Technology and Tools; Pirnay-Dummer, Ifenthaler, & Spector, 2010), which combines qualitative and quantitative research methods and provides bridges between them.

In our current research we are investigating effective instructional interventions for self-regulated learning within problem-solving processes (e.g. Ifenthaler, 2009; Ifenthaler, Masduki, & Seel, 2011). Hence, the present study was conducted to explore and evaluate different types of prompts for self-regulated learning in a problem-solving scenario. Furthermore, we introduce an automated assessment and analysis methodology for determining the quality of learning outcomes.

Cognitive processes and problem solving

A central assumption of cognitive psychology is that mental representations enable individuals to understand and explain experience and events, process information, and solve problems (Johnson-Laird, 1989). More specifically, Rumelhart, Smolensky, McClelland, and Hinton (1986) argue that these internal functions of the human mind are dependent on two interacting modules or sets of units: Schemata and mental models. In this context, schemata and mental models are theoretical constructs which specify different functions of human information processing. The resulting cognitive architecture corresponds to a great extent to Piaget’s epistemology (1943, 1976) and its basic mechanisms of assimilation and accommodation.

Accordingly, assimilation is dependent on the availability and activation of schemata, which allow new information to be integrated immediately into pre-existing cognitive structures. As soon as a schema can be activated, it runs automatically and regulates information processing. If a schema does not fit immediately into the requirements of a new problem-solving task it can be adjusted to meet them by means of accretion, tuning, or reorganization (Seel, et al., 2009). Accordingly, if a schema for any problem type is available, it is promptly mapped onto the problem to be solved (Jonassen, 2000). If assimilation is not successful, accommodation must take place in order to reorganize or restructure an individual’s knowledge. However, when no schema is available at all or when its reorganization fails, the human mind switches to the construction of a mental model, which is defined as a dynamic ad hoc representation of a phenomenon or problem that aims at creating subjective plausibility through the simplification or envisioning of the situation, analogical reasoning, or mental simulation.

We further argue that a learner constructs a mental model by integrating relevant bits of domain-specific knowledge into a coherent structure step by step in order to meet the requirements of a phenomenon to be explained or a problem to be solved. From an instructional point of view, providing direct or indirect external support within this step-by-step process could be an effective way to guide learners through problem-solving processes and facilitate their self-regulated learning in the long run. Winne (2001) provides an in-depth discussion on the above introduced concepts.

The role of metacognition and reflection in problem solving

Various researchers have highlighted the importance of metacognition for the adjustment and the regulation of learning and problem-solving activities (e.g., Boekaerts, 1999; Mayer, 1998; Schmidt-Weigand, Hänze, & Wodzinski, 2009; Zimmerman & Schunk, 2001). According to Pintrich (2000), metacognition is defined as a superordinate ability to direct and regulate cognitive, motivational, and behavioral learning and problem-solving processes in order to achieve a specific goal. Generally, researchers distinguish between two major components of metacognition, namely knowledge of cognition and regulation of cognition. Knowledge of cognition includes declarative knowledge about the self as a learner and problem-solving strategies, procedural knowledge about how to use these strategies, and conditional knowledge about when and why to use them – this metacognitive knowledge is also referred to as metacognitive awareness. Regulation of cognition, on the other hand, refers to components which facilitate the control and regulation of learning. These skills involve abilities such as planning, self-monitoring, and self-evaluation (Schraw & Dennison, 1994).

But how do learners transfer their knowledge of effective problem solving to regulate their problem-solving activities? In general, the key link between knowledge about and the regulation of one’s own problem-solving
activities is assumed to be reflective thinking (see Ertmer & Newby, 1996). If learners manage to generate information about the efficiency of their problem-solving strategies and successfully implement these findings in the ongoing problem-solving process, they are able to control and regulate their cognitive activities. Thus, metacognition refers to the ability to reflect on, understand, and control one’s learning and problem-solving activities (Simons, 1993). Accordingly, we have to distinguish between three different levels of learner-orientated reflective thinking: (1) a problem-based reflection of the learning content, (2) a behavior-oriented reflection of one’s own problem-solving activities, and (3) the learner’s identity-based reflection of his or her own learning ability. While the superordinate level of reflection requires the progressive verification of existing beliefs and established practices of one’s own learning, the behavior-oriented reflection takes place in the wake of experience (see Jenert, 2008). Furthermore, according to Wirth (2009, p. 91), “teaching learning regulation means to regulate the learner’s learning regulation.” This leads to the question of how to support learners’ reflection through instruction.

Supporting learners’ reflection via prompting

The instructional goal of teaching self-regulated problem solving is a highly demanding task (Wirth, 2009). It requires supporting learners in the acquisition and application of strategic knowledge for effective problem solving. The self-regulated learner possesses a set of problem-solving strategies and most importantly the ability to transfer and to apply this knowledge to different problem situations. In the course of their development from novice to expert, learners need guidance to learn how to regulate their problem-solving activities. Accordingly, the type of instructional aid depends on the state of the learner (grade of self-regulation). Novice learners (in terms of their self-regulation abilities) may need stronger guidance whereas expert learners do need less or no guidance at all. Hence, this decrease of strength in guidance could be described as fading of guidance (Collins, et al., 1989). On the other hand, learners also need a certain extent of autonomy to self-regulate their problem-solving activities in terms of learning by doing. The problem of accomplishing a balance between support and autonomy is referred to as the “assistance dilemma” (Koedinger & Aleven, 2007, p. 239). Additionally, in order to provide an optimal balance between external assistance and the facilitation of autonomous learning, it is necessary to distinguish between ability deficiency and production deficiency (Veenman, Kerseboom, & Imthorn, 2000). Learners with an ability deficiency suffer from a lack of metacognitive knowledge and skills. Accordingly, teachers have to convey problem-solving strategies to the learners and provide them with opportunities to exercise and reflect on their knowledge. In the case of a production deficiency, learners actually possess the knowledge and skills to regulate their problem-solving processes. However, they fail to use the inert knowledge and skills in specific problem-solving situations. In such cases, instructional support can be reduced to the activation of knowledge and skills in order to not restrict the learners in their autonomy.

Prompting is an instructional method for guiding and supporting the regulation of the learner’s problem solving processes. Prompts are presented as simple questions (e.g., “What will be your first step when solving the problem?”), incomplete sentences (e.g., “To approach the solution to the problem step by step, I have to …”), execution instructions (e.g., “First, draw the most important concepts and link them.”), or pictures and graphics (Bannert, 2007, 2009). The main goal of the method is to focus the learner’s attention on specific aspects of his or her own problem-solving process. By activating learners and motivating them to think about the efficiency of their strategies, one can increase their awareness for mostly unconsidered problem-solving activities. Therefore, they reflect on their own thoughts and are able to monitor, control, and regulate their strategic procedure in a specific situation (see Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & Lavancher, 1994; Davis, 2003; Davis & Linn, 2000; Ertmer & Newby, 1996; Ge & Land, 2004; Lin & Lehmann, 1999). The best point in time to present a prompt depends on the intention of the specific intervention. Learners should receive the prompt just in time, i.e. at the moment in which they require external support. Otherwise, these short interventions might result in cognitive overload (Thillmann, Künsting, Wirth, & Leutner, 2009). In general, a distinction is made between presentation before, during, or after a learning sequence. If the prompt is intended to activate the learners to monitor their problem-solving activities, presentation during a learning sequence is reasonable. If the intention is to induce the learners to assess certain problem-solving activities, presentation after the sequence is appropriate. Presenting the prompt before a problem-solving sequence is expedient when one wishes to inspire the learners to generate an approach to the problem-solving situation (Davis, 2003). Another crucial aspect is how metacognitive prompts can be designed and embedded to provide an optimal scaffold to the learners. Davis (2003) investigated the efficiency of reflective prompts and differentiates between generic and directed prompts. Her primary interest was to explore whether learners merely need to be prompted to reflect or need more guidance in order to reflect productively.
Accordingly, the presentation of generic prompts would seem to be more effective, because the learner’s autonomy is not undermined. The directed prompt, on the other hand, additionally asks learners to process more information, because it introduces a new expert model for reflection (see Davis, 2003).

To sum up, prompting is an instructional method that guides learners during problem-solving processes. Well-designed and embedded prompts may direct learners to perform a specific desired activity, which is contextualized within a particular problem-solving situation. Accordingly, more empirical evidence is needed to investigate which type of prompting leads to a better performance (generic vs. directed; see Davis, 2003).

New ways of assessment and analysis

Cognitive and educational researchers use theoretical constructs, e.g. metacognition, mental models, schemata, etc., to explain complex cognitive structures and procedures for learning, reasoning, and problem solving (Seel, et al., 2009). However, these internal cognitive structures and functions are not directly observable, which leads to biased assessment and analysis. Accordingly, the assessment and analysis of internal cognitive structures and functions requires that they be externalized. Therefore, we argue that it is essential to identify economic, fast, reliable, and valid techniques to elicit and analyze these cognitive structures (see Ifenthaler, 2008, 2010b). Appropriate standard methodologies include standardized questionnaires and interviews (Zimmerman, 2008), think-aloud protocols (Ericsson & Simon, 1993), the assessment of log files or click streams (Chung & Baker, 2003; Dummer & Ifenthaler, 2005; Veenman, Wilhelm, & Beishuizen, 2004), and eye-tracking measures (Mikkilä-Erdmann, Penttinen, Anto, & Ollkunoua, 2008) as well as mind tools (Jonassen & Cho, 2008). However, the possibilities of externalization are limited to a few sets of sign and symbol systems (Seel, 1999b) – characterized as graphical- and language-based approaches (Ifenthaler, 2010b). A widely accepted application is concept, causal, or knowledge maps which are automatically scored and compared to an expert’s solution (Herl, Baker, & Niemi, 1996; Ifenthaler, 2010a).

However, current discussion about the above-described methodological options suggests that it will be necessary to find new assessment and analysis alternatives (Ifenthaler, 2008; Seel, 1999a; Veenman, 2007; Veenman, et al., 2006). As not every available methodology is suitable for this research, we have introduced our own web-based assessment and analysis platform, HIMATT (Highly Integrated Model Assessment Technology and Tools; Pirnay-Dummer, et al., 2010).

HIMATT is a combined toolset which was developed to convey the benefits of various methodological approaches in a single environment and which can be used by researchers with only little prior training (Pirnay-Dummer & Ifenthaler, 2010). Methodologically, the tools integrated into HIMATT touch the boundaries of qualitative and quantitative research methods and provide bridges between them. First of all, text can be analyzed very quickly without loosening the associative strength of natural language. Furthermore, concept maps can be annotated by experts and compared to other solutions. The automated analysis function produces measures which range from surface-oriented structural comparisons to integrated semantic similarity measures. There are four structural (surface, graphical, structural, and gamma matching) and three semantic (concept, propositional, and balanced propositional matching) measures available (see the Method section for a detailed description of them). All of the data, regardless of how it is assessed, can be analyzed quantitatively with the same comparison functions for all built-in tools without further manual effort or recoding. Additionally, HIMATT generates standardized images of text and graphical representations (Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer, et al., 2010).

Research questions and hypotheses

The central research objective of this study is to identify the efficiency of different types of prompts (generic vs. directed) for activating learners to reflect on their ongoing problem-solving process. Based on prior research (Davis, 2003; Ge & Land, 2004), we hypothesized that learners who receive generic prompts during the problem-solving process will perform better than those who receive directed prompts. Accordingly, a generic prompt provides learners necessary support and allows them a certain extent of autonomy to self-regulate their problem-solving activities (Koedinger & Aleven, 2007). Hence, we assume that learners who receive generic prompts will perform better with regard to their domain-specific understanding (Hypothesis 1). If learners do not already possess the required self-regulative knowledge and skills, directed prompts would be more effective. Additionally, we assume
that the problem representations (in the form of a concept map) of learners with generic prompts will be structurally (Hypothesis 2) and semantically (Hypothesis 3) more similar to an expert’s solution than those of learners who have received directed prompts.

Additionally, previous research studies have found contradictory results concerning learners’ metacognitive processes and deductive reasoning skills in association with learning outcomes when working with concept maps in problem solving scenarios (e.g. Hilbert & Renkl, 2008; Ifenthaler, Pirnay-Dummer, & Seel, 2007; O'Donnell, Dansereau, & Hall, 2002; Veenman, et al., 2004). We assume that learners with higher metacognitive awareness will outperform those with lower metacognitive awareness with regard to their learning outcomes (Hypothesis 4a). Additionally, we assume that better deductive reasoning skills will have a positive effect on the learning outcomes (Hypothesis 4b).

Method

Participants

Ninety-eight students (68 female and 30 male) from a European university participated in the study. Their average age was 21.9 years (SD = 3.5). They were all enrolled in an introductory course on research methods and had studied for an average of 2.4 semesters (SD = 3.1).

Design

Participants were randomly assigned to the three experimental conditions. The three experimental conditions were related to the three forms of reflective thinking prompt: generic prompt (GP; n1 = 32), direct prompt (DP; n2 = 40), and control group (CG; n3 = 26). Participants in the GP group received general instructions for planning and reflecting on their ongoing problem-solving activities (see materials for details). For participants in the DP group, we provided nine sentences which referred to planning (1–3), monitoring (4–6), and evaluation (7–9) of the ongoing problem-solving activities (see materials for details). The CG did not receive a reflective thinking prompt. ANOVA was used to test for study experience differences (number of semesters studied) among the three experimental groups. The experimental groups did not differ with regard to the semesters studied, F(2, 95) = 0.42, p > .05.

Materials

Problem scenario

A German-language article on the human immune system and the consequences of virus infections with 1,120 words was used as learning content. The problem was to identify differences between an influenza and HIV infection. Specifically, the problem task consisted of the following two questions: (1) What happens to the immune system during an initial infection with the influenza virus? (2) What effect does an HIV infection have on the immune system in contrast to an influenza infection? Additionally, learners were asked to graphically represent their understanding of these complex biological processes (questions one and two) in form of a concept map. Also, an expert solution (based on the article) in the form of a concept map was generated which functioned as a reference model for later analysis.

Domain specific knowledge test

The knowledge test included 13 multiple-choice questions with four possible solutions each (1 correct, 3 incorrect). First, 20 questions were developed on the basis of the article on the human immune system and the consequences of virus infections. Second, in a pilot study (N = 10 participants), we tested the average difficulty level to account for ceiling effects. Finally, we excluded seven questions because they were not appropriate for our experimental study. In our experiment we administered two versions (in which the 13 multiple-choice questions appeared in a different order) of the domain-specific knowledge test (pre- and posttest). It took about eight minutes to complete the test.
Metacognitive awareness inventory

The participants’ metacognitive awareness was assessed with the Metacognitive Awareness Inventory (Schraw & Dennison, 1994). Each of the 52 items of the inventory was answered on a scale from 1 to 100 (Cronbach’s alpha = .90). Two dimensions of metacognitive awareness were addressed: (1) knowledge of cognition, which includes knowledge about personal skills, learning strategies, and the efficiency of these strategies, and (2) regulation of cognition, which includes planning and initiating of learning, implementation of strategies, monitoring and control of learning, and the evaluation of personal learning efficiency.

Deductive reasoning inventory

A subscale of the ASK (Analyse des Schlussfolgernden und Kreativen Denkens; i.e. inventory for deductive reasoning and creative thinking) was used to test the participants’ deductive reasoning (Schuler & Hell, 2005). The subscale included questions on the interpretation of information (21 items), drawing conclusions (32 items), and facts and opinions (27 items). Schuler and Hell (2005) report good reliability scores for the ASK (Cronbach’s alpha = .72; test-retest reliability = .78).

Experience with concept mapping test

The participants’ experience with concept mapping was tested with a questionnaire including eight items (Ifenthaler, 2009; Cronbach’s alpha = .87). The questions were answered on a five-point Likert scale (1 = totally disagree; 2 = disagree; 3 = partially agree; 4 = agree; 5 = totally agree). Items included in the test, e.g. “I use concept maps to structure learning content”, “The construction of a concept map raises no difficulties”, or “I use computer software for constructing concept maps” (translated from German).

Reflective thinking prompts

Two versions of prompts were developed in order to stimulate the participants to reflect on their problem-solving activities. (1) The generic prompt (“stop and reflect”) included the following advice: “Use the next 15 minutes for reflection. Reflect critically on the course and outcome of your problem-solving process. Amend and improve your concept map if necessary. Feel free to use all materials provided! (translated from German).” (2) The direct prompt included the following advice: “Use the next 15 minutes for reflection. Reflect critically on the course and outcome of your problem-solving process. Feel free to use all materials provided! The guidelines provided below may be used as an aid. Please complete the list item by item by completing each sentence on its own in your mind. 1. The requirements/goals of the problem included...; 2. The basic conditions which had to be taking in account to complete this problem were...; 3. In order to find the best solution to the problem, I...; 4. In order to understand the context and main ideas of the text, I...; 5. In order to come a bit closer to the solution with each step, I...; 6. In order to create an optimal concept map of the text, I...; 7. I believe I solved the problem well, because...; 8. I could solve the problem better next time if I...; 9. In order to improve my explanation model I will now... (translated from German).”

HIMATT concept mapping tool

The concept mapping tool, which is part of the HIMATT (Pirmay-Dummer, et al., 2010) environment, was used to assess the participants’ understanding of the problem scenario. The intuitive web-based tool allows participants to create concept maps with only little training (Pirmay-Dummer & Ifenthaler, 2010). Once created, all concept maps are automatically stored on the HIMATT database for further analysis.

Procedure

First, the participants were randomly assigned to the three experimental conditions (GP, DP, CG). Then they completed a demographic data survey (three minutes), the metacognitive awareness inventory (ten minutes), the
deductive reasoning inventory (33 minutes), and the experience with concept mapping test (five minutes). Next, the participants were given an introduction to concept maps and were shown how to use the HIMATT environment (ten minutes). After a short relaxation phase (five minutes), they answered the 13 multiple choice questions of the domain-specific knowledge test on the immune system and the consequences of virus infections (pretest; eight minutes). Then they received the article on the immune system and the consequences of virus infections and were introduced into the problem scenario. In total, all participants spent 25 minutes on the problem scenario. Additionally, participants in the experimental condition GP and DP received their reflective thinking prompt after 15 minutes working on the problem scenario. The CG did not receive a reflective thinking prompt. They were allowed to take notes with paper and pencil. After another short relaxation phase (five minutes), the participants logged into the HIMATT environment and constructed a concept map on their understanding of the problem scenario (ten minutes). Finally, the participants answered the 13 multiple choice questions of the posttest on declarative knowledge (eight minutes).

Data analysis

In order to analyze the participants’ understanding of the problem scenario, we used the seven measures implemented in HIMATT (see Ifenthaler, 2010b; Pirnay-Dummer, et al., 2010). Accordingly, each of the participants’ concept maps was compared automatically against the reference map (expert solution based on the article). Table 1 describes the seven measures of HIMATT, which include four structural measures and three semantic measures (Ifenthaler, 2010a, 2010b; Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer, et al., 2010). HIMATT uses specific automated comparison algorithms to calculate similarities between a given pair of frequencies $f_1$ (e.g. expert solution) and $f_2$ (e.g. participant solution). The similarity $s$ is generally derived by

$$s = 1 - \frac{|f_1 - f_2|}{\max\{f_1, f_2\}}$$

which results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is identity. The other measures collect sets of properties. In this case, the Tversky similarity (Tversky, 1977) applies for the given sets A (e.g. expert solution) and B (e.g. participant solution):

$$s = \frac{f(A \cap B)}{f(A \cap B) + \alpha \cdot f(A - B) + \beta \cdot f(B - A)}$$

$\alpha$ and $\beta$ are weights for the difference quantities which separate $A$ and $B$. They are usually equal ($\alpha = \beta = 0.5$) when the sources of data are equal. However, they can be used to balance different sources systematically, e.g. comparing a learner’s concept map which was constructed within five minutes to an expert’s concept map, which may be an illustration of the result of a conference or of a whole book (see Pirnay-Dummer & Ifenthaler, 2010). The Tversky similarity also results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is identity.

Reliability scores exist for the single measures integrated into HIMATT. They range from $r = .79$ to $r = .94$ and are tested for the semantic and structural measures separately and across different knowledge domains (Pirnay-Dummer, et al., 2010). Validity scores are also reported separately for the structural and semantic measures. Convergent validity 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).

<table>
<thead>
<tr>
<th>Measure [abbreviation] and type</th>
<th>Short description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface matching [SFM] Structural indicator</td>
<td>The surface matching (Ifenthaler, 2010a) compares the number of vertices within two graphs. It is a simple and easy way to calculate values for surface complexity.</td>
</tr>
<tr>
<td>Graphical matching [GRM] Structural indicator</td>
<td>The graphical matching (Ifenthaler, 2010a) compares the diameters of the spanning trees of the graphs, which is an indicator for the range of conceptual knowledge. It corresponds to structural matching as it is also a measure for structural complexity only.</td>
</tr>
</tbody>
</table>
### 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 $d$ (small effect: $d < .50$, medium effect $.50 \leq d \leq .80$, strong effect $d > .80$) and partial $\eta^2$ (small effect: $\eta^2 < .06$, medium effect $.06 \leq \eta^2 \leq .13$, strong effect $\eta^2 > .13$).

More than half of the participants (58%) did not use concept maps to structure their own learning materials before our experiment. Only 5% of the participants used concept mapping software to create their own concept maps beforehand. On the other hand, over 60% of the participants answered that they did not find it difficult to create a concept map. Consequently, there was no significant difference in the learning outcome as measured by the domain-specific knowledge posttest between participants who used concept mapping software before the experiment and those who did not use concept mapping software at all, $t(96) = .105, ns$.

### Domain-specific knowledge

On the domain-specific knowledge test (pre- and posttest), participants could score a maximum of 13 correct answers. In the pretest they scored an average of $M = 4.38$ correct answers ($SD = 1.71$) and in the posttest $M = 6.71$ correct answers ($SD = 2.49$). The increase in correct answers was significant, $t(97) = 9.611, p < .001, d = 1.068$. ANOVA was used to test for knowledge gain differences among the three experimental groups. The experimental groups did not differ with regard to the results in the pretest, $F(2, 95) = 2.14, p > .05$. However, the increase in correct answers differed significantly across the three experimental groups, $F(2, 95) = 8.21, p = .001, \eta^2 = .147$. Tukey HSD post-hoc comparisons of the three groups indicate that the generic prompt group ($M = 3.66, SD = 2.40, 95\% CI [2.79, 4.52]$) gained significantly more correct answers than the directed prompt group ($M = 1.68, SD = 2.14, 95\% CI [.99, 2.36]$), $p = .001$, and the control group ($M = 1.73, SD = 2.20, 95\% CI [.84, 2.62]$), $p = .005$. Comparisons between the directed prompt group and the control group were not statistically significant at $p < .05$. Accordingly, the results support the hypothesis that participants who receive generic prompts outperform those in other groups with regard to their domain-specific understanding.
**HIMATT structural measures**

The participants’ understanding of the problem scenario as illustrated by concept maps was analyzed automatically with the HIMATT tool. The four structural measures reported in Table 2 show the average similarity between the participants’ solution and the referent solution (expert concept map). Four separate ANOVAs (for HIMATT measures SFM, GRM, STM, GAM) with Tukey HSD post-hoc comparisons were computed to test for differences between the three experimental groups.

ANOVA revealed a significant difference between participants in the three experimental groups for the HIMATT measure STM, \( F(2, 95) = 7.77, p = .001, \eta^2 = .141 \). Tukey HSD post-hoc comparisons of the three groups indicate that the complete structure (STM) of the generic prompt group’s concept maps (\( M = .84, SD = .14, 95\% CI [.79, .89] \)) was significantly more similar to the expert solution than that of the directed prompt group’s maps (\( M = .70, SD = .14, 95\% CI [.66, .75], p = .001 \)). Additionally, the complete structure (STM) of the control group’s concept maps (\( M = .80, SD = .19, 95\% CI [.73, .88] \)) was significantly more similar to the expert solution than that of the directed prompt group’s maps, \( p = .026 \). Comparisons between the directed prompt group and the control group were not statistically significant at \( p < .05 \).

For the HIMATT measure GAM, ANOVA revealed a significant difference between the three experimental groups, \( F(2, 95) = 5.49, p = .006, \eta^2 = .104 \). Tukey HSD post-hoc comparisons of the three groups indicate that the density of vertices (GAM) of the generic prompt group’s concept maps (\( M = .83, SD = .10, 95\% CI [.79, .87] \)) was significantly more similar to the expert solution than that of the directed prompt group’s maps (\( M = .70, SD = .19, 95\% CI [.64, .76], p = .004 \)). All other comparisons between groups were not statistically significant at \( p < .05 \).

ANOVA for the HIMATT measures SFM and GRM revealed no significant differences between the experimental groups. Accordingly, the results support the hypothesis that participants who receive generic prompts outperform participants in other groups with regard to the HIMATT measures STM and GAM.

**Table 2. Means (SD) HIMATT structural measures for the three experimental groups \( (N = 98) \)**

<table>
<thead>
<tr>
<th></th>
<th>GP ( (n_1 = 32) )</th>
<th>DP ( (n_2 = 40) )</th>
<th>CG ( (n_3 = 26) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface matching [SFM]</td>
<td>.73 (.19)</td>
<td>.60 (.25)</td>
<td>.68 (.28)</td>
</tr>
<tr>
<td>Graphical matching [GRM]</td>
<td>.77 (.18)</td>
<td>.72 (.21)</td>
<td>.71 (.21)</td>
</tr>
<tr>
<td>Structural matching [STM]</td>
<td>.84 (.14)</td>
<td>.70 (.14)</td>
<td>.80 (.19)</td>
</tr>
<tr>
<td>Gamma matching [GAM]</td>
<td>.83 (.10)</td>
<td>.70 (.19)</td>
<td>.74 (.19)</td>
</tr>
</tbody>
</table>

*Note.* HIMATT similarity measures between participant’s solution and expert’s solution (0 = no similarity; 1 = total similarity); GP = generic prompt, DP = directed prompt, CG = control group

**HIMATT semantic measures**

Additional HIMATT analysis for the semantic measures of the participants’ understanding of the problem scenario as expressed by concept maps was computed. The three semantic measures reported in Table 3 show the average similarity between the participants’ solution and the referent solution (expert concept map). Three separate ANOVAs (for HIMATT measures CCM, PPM, BPM) with Tukey HSD post-hoc comparisons were computed to test for differences between the three experimental groups.

ANOVA revealed a significant difference between participants in the three experimental groups for the HIMATT measure CCM, \( F(2, 95) = 7.40, p = .001, \eta^2 = .135 \). Tukey HSD post-hoc comparisons of the three groups indicate that the semantic correctness of single concepts used in the concept maps (CCM) of the generic prompt group (\( M = .43, SD = .19, 95\% CI [.37, .50] \)) was significantly more similar to the expert solution than in those of the directed prompt group (\( M = .30, SD = .14, 95\% CI [.26, .34], p = .001 \)), and the control group (\( M = .31, SD = .15, 95\% CI [.25, .37], p = .011 \)). Comparisons between the directed prompt group and the control group were not statistically significant at \( p < .05 \).

ANOVA revealed a significant difference between participants in the three experimental groups for the HIMATT measure PPM, \( F(2, 95) = 10.80, p < .001, \eta^2 = .185 \). Tukey HSD post-hoc comparisons of the three groups indicate that the semantic correctness of propositions (concept-link-concept) used in the concept maps (PPM) of the generic prompt group (\( M = .17, SD = .16, 95\% CI [.11, .23] \)) was significantly more similar to the expert solution than in
those of the directed prompt group ($M = .06, SD = .06, 95\% CI [.04, .08]), $p < .001$, and the control group ($M = .07, SD = .08, 95\% CI [.04, .10]), $p = .002$. Comparisons between the directed prompt group and the control group were not statistically significant at $p < .05$.

Table 3. Means (SD) HIMATT semantic measures for the three experimental groups ($N = 98$)

<table>
<thead>
<tr>
<th></th>
<th>GP ($n_1 = 32$)</th>
<th>DP ($n_2 = 40$)</th>
<th>CG ($n_3 = 26$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept matching [CCM]</td>
<td>.43 (.19)</td>
<td>.30 (.14)</td>
<td>.31 (.15)</td>
</tr>
<tr>
<td>Propositional matching [PPM]</td>
<td>.17 (.16)</td>
<td>.06 (.06)</td>
<td>.07 (.08)</td>
</tr>
<tr>
<td>Balanced propositional matching [BPM]</td>
<td>.33 (.22)</td>
<td>.16 (.16)</td>
<td>.17 (.17)</td>
</tr>
</tbody>
</table>

Note. HIMATT similarity measures between participant’s solution and expert’s solution (0 = no similarity; 1 = total similarity); GP = generic prompt, DP = directed prompt, CG = control group

ANOVA revealed a significant effect between participants in the three experimental groups for the HIMATT measure BPM, $F(2, 95) = 8.97$, $p < .001$, $\eta^2 = .159$. Tukey HSD post-hoc comparisons of the three groups indicate that the quotient of the semantic correctness of propositions (concept-link-concept) and single concepts used in the concept maps (BPM) of the generic prompt group ($M = .43, SD = .19, 95\% CI [.25, .41]$) was significantly more similar to the expert solution than in those of the directed prompt group ($M = .30, SD = .14, 95\% CI [.11, .21]), $p < .001$, and the control group ($M = .31, SD = .15, 95\% CI [.11, .24]), $p = .004$. Comparisons between the directed prompt group and the control group were not statistically significant at $p < .05$. Accordingly, the results support the hypothesis that participants who receive generic prompts outperform participants in other groups with regard to the HIMATT measures CCM, PPM, and BPM.

Correlational analyses

Correlations were calculated between metacognitive awareness, deductive reasoning and the seven HIMATT measures as well as for the domain-specific knowledge of the posttest (see Table 4).

Table 4. Correlations between metacognitive awareness, deductive reasoning and HIMATT measures, domain specific knowledge (post-test)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of cognition</td>
<td>.50</td>
<td>.02</td>
<td>.163</td>
<td>.104</td>
<td>.109</td>
<td>.114</td>
<td>.178</td>
<td>-.001</td>
</tr>
<tr>
<td>Regulation of cognition</td>
<td>.88</td>
<td>.50</td>
<td>.049</td>
<td>-.051</td>
<td>.037</td>
<td>.124</td>
<td>.177</td>
<td>.076</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deductive reasoning</th>
<th>Interpretation of information</th>
<th>Drawing conclusions</th>
<th>Facts and opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.014</td>
<td>.018</td>
<td>.109</td>
</tr>
<tr>
<td></td>
<td>-.016</td>
<td>.029</td>
<td>.099</td>
</tr>
<tr>
<td></td>
<td>.072</td>
<td>.013</td>
<td>.066</td>
</tr>
<tr>
<td></td>
<td>-.030</td>
<td>.070</td>
<td>-.061</td>
</tr>
<tr>
<td></td>
<td>.104</td>
<td>-.118</td>
<td>-.072</td>
</tr>
<tr>
<td></td>
<td>.068</td>
<td>-.055</td>
<td>-.042</td>
</tr>
<tr>
<td></td>
<td>.055</td>
<td>-.034</td>
<td>.022</td>
</tr>
</tbody>
</table>

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

Positive deductive reasoning abilities were related to better domain-specific knowledge. Accordingly, interpretation of information correlated significantly with the learning outcomes as measured by the domain-specific knowledge test, $r = .351$, $p < .01$. Apparently the learners’ ability to interpret available information was associated positively with the domain-specific knowledge. Additionally, drawing conclusions correlated significantly with the learning outcomes, $r = .416$, $p < .01$. Hence, the learners’ logical reasoning from given information was strongly associated with the domain-specific knowledge. Furthermore, “facts and opinions” correlated significantly with the learning
outcomes, \( r = .297, \ p < .01 \). Accordingly, the learners’ ability to differentiate between facts and opinions was positively associated with the domain-specific knowledge.

However, no correlations were found between metacognitive awareness and the domain-specific knowledge. Finally, no correlations were found between the HIMATT measures and metacognitive awareness or deductive reasoning (see Table 4).

**Discussion**

The facilitation of self-regulated learning is a balancing act between external support and internal regulation. An instructional method for guiding and supporting the regulation of learners’ problem-solving processes is prompting. Prompts are presented as simple questions, incomplete sentences, explicit execution instructions, or pictures and graphics for a specific learning situation. Prompts are categorized in generic and directed forms. Generic prompts ask learners to stop and reflect about their current activities. Directed prompts additionally provide learners expert models of reflective thinking.

The aim of the present study was to explore the efficiency of different types of prompts for reflection in a self-regulated problem-solving situation. It was assumed that well-designed and embedded prompts may direct learners to perform successfully within a particular self-regulated problem-solving situation (Davis, 2003; Thillmann, et al., 2009). The problem was to identify differences between an influenza and HIV infection as well as their effects to the human immune system. In order to assess the participants’ understanding of the problem scenario, we asked them to create a concept map on their subjectively plausible understanding of the phenomenon in question. Three experimental conditions with different reflective thinking prompts were realized. Participants in the generic prompt group (GP) received general instructions for planning and reflecting on their ongoing problem-solving activities. For participants in the direct prompt group (DP), we provided nine sentences which referred to planning, monitoring, and evaluation of the ongoing problem-solving activities. Participants in the control group (CG) did not receive a reflective thinking prompt.

In order to analyze the elicitation of the participants’ understanding of the problem scenario, we introduced our own web-based platform HIMATT (Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer, et al., 2010). Within HIMATT, participants’ concept maps can be automatically compared to a referent map created by an expert based on the problem scenario. The HIMATT analysis function produces measures which range from surface-oriented structural comparisons to integrated semantic similarity measures. Four structural measures (surface [SFM], graphical [GRM], structural [STM], and gamma [GAM]) and three semantic measures (concept [CCM], propositional [PPM], balanced propositional [BPM]) were used to answer our research questions.

Major findings of the present study are that participants in the generic prompt group outperformed other learners with regard to their (1) domain-specific knowledge gain as well as their (2) structural and (3) semantic understanding of the problem scenario.

First, findings on domain-specific knowledge suggest that generic prompts (e.g., What will be your first step when solving the problem?) are most effective in self-regulated learning environments. Generic prompts guide learners to use a specific set of problem-solving strategies and at the same time give them a certain extent of autonomy to self-regulate their problem-solving activities (Koedinger & Aleven, 2007). In contrast, direct prompts seem to prevent learners from solving a problem autonomously. However, we believe that direct prompts could be helpful for novices who do not yet possess the necessary problem-solving skills. Hence, further empirical investigations are necessary to answer these assumptions.

Second, generic prompts also had a positive effect on the structural similarity of learners’ understanding of the problem scenario with regard to the expert solution. Compared to the expert solution, GP learners’ solutions represented more strongly connected knowledge, which could indicate a deeper subjective understanding of the underlying subject matter (HIMATT measures STM, GAM). However, the number of concepts and links (SFM) and the overall complexity of the problem representations were not influenced by the different prompts. We believe that an effect towards complexity will occur in longer perspectives requiring an in-depth analysis of the learning-dependent change (Ifenthaler, et al., 2011; Ifenthaler & Seel, 2005).
Third, findings for the semantic HIMATT measures (CCM, PPM, BPM) are in line with the above-discussed results. Solutions of GP learners are semantically more similar to the expert solution than those of other learners. However, the overall similarity of the learners’ problem representation to the expert representation is low. Hence, the learners of the present study are far from being experts and should be given more time and resources to improve their overall performance. Accordingly, we believe that further studies are needed to better understand the underlying cognitive processes of learning-dependent progression from novice to expert and, as a consequence, to provide more effective instructional materials.

Furthermore, correlational analysis showed that metacognitive awareness and deductive reasoning skills were not associated with the problem scenario representation as expressed by concept maps. These results complement previous research studies which have found similar results (e.g. Hilbert & Renkl, 2008; Ifenthaler, et al., 2007; O'Donnell, et al., 2002). However, we found significant correlations between domain-specific knowledge and deductive reasoning skills. Accordingly, deductive reasoning skills have positive effects on the declarative learning outcomes. One final consideration based on our findings is that when we train novices to become experts, we often think about training general abilities to efficiently facilitate the process. While this works well for training abilities themselves, these methods may have limits when we train experts who have to decide and act within complex domains (Chi & Glaser, 1985; Ifenthaler, 2009).

Reviewing the results of our experimental investigation, we suggest that a generic prompt which includes general instructions for planning and reflecting on their ongoing problem-solving activities are most effective for learners which already have a solid set of skills (in our case students at a university). In contrast, if learners do not have a specific set of problem-solving skills, directed prompts may be more effective. Accordingly, future studies should focus on the effectiveness of different prompts for different types of learners (novices, advanced learners, expert learners).

The present research is limited to the single problem scenario on differences between an influenza and HIV infection as well as their effects to the human immune system. The limited time and resources for solving the problem may have also had an influence on our results. Also, further empirical investigations should focus on the “best” point in time when to present a prompt (Thillmann, et al., 2009). In addition, the present research is limited by our use of concept maps to elicit the problem scenario. However, such graphical representations are a widely accepted method for illustrating the meaning of locally discussed information (Eliaa, Gagatsisa, & Demetriou, 2007; Hardy & Stadelhofer, 2006; Ruiz-Primo, Schultz, Li, & Shavelson, 2001). In order to improve the external validity of our research, we suggest applying additional methodologies such as think-aloud protocols (Ericsson & Simon, 1993), standardized questionnaires and interviews (Zimmerman, 2008), and log files or click streams (Chung & Baker, 2003; Veenman, et al., 2004) within multimedia learning environments. Especially thinking aloud protocols applied during the reflection phase could give more insights into the metacognitive procedures induced by different types of prompts. Lastly, the timing of the prompts should be investigated in future studies (Thillmann, et al., 2009). Accordingly, future studies will include not only prompts for reflecting on the problem-solving process but also reflection prompts provided before the learners enter the problem scenario.

Conclusions

To sum up, since cognitive and educational researchers are not able to measure internal cognitive structures and functions directly, studies like ours will always be biased. A major bias includes the limited possibilities for externalizing learners’ internal cognitive structures (Ifenthaler, 2008, 2010b). However, we are adamant in our belief that it is essential to identify economic, fast, reliable, and valid methodologies to elicit and analyze these cognitive structures and functions (Zimmerman, 2008). In conclusion, new ways of assessment and analysis could make more precise results available, which may in turn lead to superior instructional interventions in the future.

References


Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science, 39*(1), 41-61. doi: 10.1007/s11251-009-9097-6


ABSTRACT
Technology offers many opportunities for educators to support teaching, learning and assessment. This paper introduces a project to design and implement a virtual environment (SAVE Science) intended to assess (not teach) middle school students’ knowledge and use of scientific inquiry through two modules developed around curriculum taught in middle schools in Pennsylvania, U.S.A. We explore how the concept of ‘presence’ supports these efforts, as well as how Piaget’s theory of developmental stages can be used as a lens to understand whether these students achieved presence in the modules. Findings are presented from a study looking at 154 middle school students’ perceived sense of presence in a virtual world developed for the SAVE Science research project as demonstrated through a post module survey and a post module discussion with their teacher. Age and gender differences are explored. In addition we use content analysis, as described by Slater and Usoh (1993), of student talk in the post module discussion transcripts to identify levels of “presence.” In the end, participating seventh grade students demonstrated achieving some level of presence, while the sixth grade students did not.

Keywords
Presence, Middle school, Immersive virtual environments, Piaget

Introduction

As we move further into the twenty-first century, technology opportunities allow researchers to consider many ways in which technology can be exploited to support teaching, learning, and assessment of learning. (Gee, 2003) Most research on educational uses of technologies focuses on how they are used to teach, rather than on how they are used to assess knowledge or application of knowledge. The Situated Assessment using Virtual Environments for Science Inquiry and Content (SAVE Science) research project is one that is focused on assessment rather than learning. The SAVE Science project is designing and implementing a series of virtual environment-based assessments of middle school science content. Students engage individually with a contextualized problem and use knowledge learned in the classroom to solve it. However, in order for students to fully engage in the tasks presented in the SAVE Science modules, we hypothesize that the more they perceive themselves as part of the story, the more likely they will actively engage in the activities. Thus, the students need to perceive they are “present” in the story/module.

The concept of ‘presence’ has been the subject of research and discussion for over 30 years. For many scholars, presence has its roots in symbolic interactionism, or how we make meaning of new experiences based on prior experiences with similar events/items/ideas (Mead & Morris, 1934, Blumer, 1969), and in social psychology theories of interpersonal communication, as in Goffman’s concept of copresence, or how we make meaning of new experiences based on prior interaction. Research on the concept of presence has included all types of media, including letters, television, telephones, teleconferencing systems, immersive VE, and virtual games (Daft & Lengel, 1984, Lombard & Ditton, 1997). Steuer (1993) suggests the relationship between vividness and interactivity on a graph (interactivity on the x-axis and vividness on the y-axis) with the book being low on both scales, Star Trek’s Holodeck as being high on both scales, and a range of other technologies located in the scatter plot in between (p. 90). With the advent of commercial Multi-User Virtual Environments such as World of Warcraft and Second Life, Delgarno and Lee (2010) suggest that presence be defined as “being there together.” But what is ‘presence’?
Initially this paper explores definitions of presence put forth over the last 20 years, and how these apply to middle school students rather than adults. Second, we introduce the reader to the SAVE Science project. Third, we present preliminary findings from a SAVE Science study exploring whether participating students achieved presence through the VE modules as demonstrated through an online survey and a post-module discussion with their teacher, and if there is a difference in perceived presence based on gender or age, using Piaget’s levels of development as a lens.

**Defining presence within virtual environments**

Historically, presence has been discussed in terms of how an individual’s interactions with a virtual reality are depicted in the virtual environment (VE). The key factors considered important to presence were sensory, perceptual, and social (see definitions below). The type/design of the environment was extremely important to facilitate development of presence or not. Table 1 summarizes several papers on these factors. Heeter (1992), in addressing Virtual Reality (VR) technologies, used the term “dimensions” to define presence and considered the personal (or how much you perceived/believed you were part of the VR), social (or how much you perceived/believed “others” in the VR interacted with you), and environmental (or how much you perceived/believed the environment reacted to you as part of the scene) aspects of interaction in the VR. Steuer (1993), discussing how VR is technologically rather than experientially focused, suggested two dimensions of presence: vividness and interactivity. The more vivid the VR environment and the more one could interact with the environment and characters, the more likely one would experience the sensation of being in that environment. One could argue that these claims are complementary, because in order to perceive/believe you are “in” a VR, you need to sense (see, hear, feel) that environment in a way that is real, and you need interact with characters and objects found within that environment. But being in a VR is very different today from what it was in the early 1990s.

Lombard and Ditton (1997) chose to define presence by considering three causes of it. First, they discussed form variables based on sensory richness, as defined by Steuer (1995). The more senses that were induced to participate, the more likely the user would achieve a sense of presence through the medium. The addition of rich natural voice and rich visual sensory stimuli produced a greater sense of presence than one or the other alone. Next were content variables where the responses to the user were perceived as being by another social entity, and not the medium or computer. The more the computer used the natural language of the users, the more likely the users would “believe” they were interacting with another social being. Lastly, they described the user variables, which included past experience with the medium, age, gender, and personality. These are ones we will use later in looking at middle school students’ reactions to SAVE Science modules.

<table>
<thead>
<tr>
<th>Author, date</th>
<th>Conditions / Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heeter, 1992</td>
<td>three dimensions of presence</td>
</tr>
<tr>
<td></td>
<td>1. personal, as you feel like you are in a VE</td>
</tr>
<tr>
<td></td>
<td>2. social, or the extent to which beings exist and react to you in the VE</td>
</tr>
<tr>
<td></td>
<td>3. environmental, as the VE reacts to you</td>
</tr>
<tr>
<td>Steuer, 1993, 1995</td>
<td>two determinants of telepresence</td>
</tr>
<tr>
<td></td>
<td>1. vividness, or the sensorially rich mediated environment, stimulus driven, and depth</td>
</tr>
<tr>
<td></td>
<td>of sensory information, dimensionality (1, 2, or 3D)</td>
</tr>
<tr>
<td></td>
<td>2. interactive, or the degree to which the user can influence form and content, number of</td>
</tr>
<tr>
<td></td>
<td>people interacting in real time</td>
</tr>
<tr>
<td>Lombard &amp; Ditton, 1997</td>
<td>three causes of presence</td>
</tr>
<tr>
<td></td>
<td>1. form variables, including interactivity, use of voice, medium and shapes</td>
</tr>
<tr>
<td></td>
<td>2. content variables, including social realism, media conventions, and nature of task</td>
</tr>
<tr>
<td></td>
<td>3. user variables, including past experience with the medium, age, gender, and personality</td>
</tr>
<tr>
<td>Witmer &amp; Singer, 1998</td>
<td>three conditions for presence</td>
</tr>
<tr>
<td></td>
<td>1. level of involvement, depends on the degree of significance given to various stimuli</td>
</tr>
<tr>
<td></td>
<td>2. ability of user to focus on the virtual world</td>
</tr>
<tr>
<td></td>
<td>3. degree of immersion, or psychological state of being included in and interacting with the VE</td>
</tr>
<tr>
<td>Delgarno &amp; Lee, 2010</td>
<td>being there together in a multi-user VE with others around the world, e.g., Second Life</td>
</tr>
</tbody>
</table>
Witmer and Singer (1998) suggested similar concepts for how presence is achieved, identifying the importance of the ability of the user to concentrate on the VE to the exclusion of all other stimuli. Schubert, Friedmann, and Regenbrecht (2001) supported this by saying, “…the sense of presence should involve two components: the sense that we are located in and acting from within the VE, and the sense that we are concentrating on the VE and ignoring the real environment.” (p. 269) As noted by Mikropoulos, “researchers agree with the description of presence as ‘the sense of being there’.” (2006, p. 197)

In reviewing MUVEs, such as World of Warcraft and Second Life where players interact with others worldwide, Delgarno and Lee (2010) explored the relationship between immersion and presence, offering a model of learning in MUVEs with distinguishing characteristics of MUVEs being representational fidelity and learner interaction (See Delgarno & Lee, 2010, p. 15 for the model). They proposed that representational fidelity, as defined in their model, incorporated aspects of both single-user and multi-user environments, including Brna’s concept of social fidelity, which combined social familiarity and social reality (Brna, 1999). From these concepts, they posited that presence in a MUVE is “being there together” because people worldwide participate. They also suggested this is true about single-user VEs.

Reviewing these various definitions, considerations and conditions, one can conclude that achieving presence in all kinds of VE depends on the ability of the user/player to seriously concentrate on what is happening in the VE, close out any competing distractions from the real world, and believe the avatar they choose/use is actually them interacting with other avatars and characters in the VE. All of this depends on the level of perceived immersion into the virtual world, and the level of interactivity made available to the user/player.

Presence, age, and gender of participant

Another concept to consider is the targeted populations in these studies. For instance, Gibson’s work studied how adult pilots land planes using landmarks to afford them a sense of distance. The articles reviewed above either did not specify the age range of their sample, or they used college-age students as the subjects (Witmer & Singer, 1998). However, in most cases gender was not identified.

How individuals react to/in a virtual world and/or perceive themselves as “being there” or “being there together” (i.e., present) in the world may also depend on the development stage of the users/players. Lombard and Ditton (1997) hinted at this concept when they suggested the “user variable” that included age of the user. They referenced Turkle (1984) who suggested that adults think of computers as machines, while children “enter into ‘social relationships’ with the machine in which they [get] competitive, angry, and even vindictive” (p. 47). Taking cues from Piaget’s theory (1953) of development of children’s understanding of their world, it seems probable that stage of development is a crucial variable in development of presence in a VE. Piaget considered children ages 7-12 to be at the “Concrete Operations” stage of development, where thoughts are more logical. At this stage, children are developing the ability to classify information, problem-solve in a concrete, systematic way, but are unable to think abstractly. Children older than 12 (through adulthood) are classified as in the “Formal Operations” stage, described as thinking more abstractly, incorporating formal logic where they could generate multiple hypotheses and possible outcomes, and thinking less tied to concrete reality.

Research studies with children as participants mainly focused on the use of digital games. (e.g. Squire, 2002; Facer, 2003; Fromme, 2003; Virvou & Katsionis, 2006; Kim, Park, & Baek, 2008; Papastergiou, 2009) However, these authors discussed the impact of serious games on student achievement without studying how the students perceived their “being there” in the virtual game environment. Squire (2002) concentrated on discussing how games are perceived in education, while unpacking gameplay through learning science, but again not discussing the ages of users. Others concentrated on such aspects as skill, communication and working with others, problem-solving, and mathematical development (Facer, p. 5), games as part of children’s culture (Fromme, 2003), usability and likability of games by children (Virvou & Katsionis, 2006), meta-cognitive strategies used by 9th graders (Kim, Park & Baek, 2008), and games in high school computer science class (Papastergiou, 2009). None of these studies considered Piaget’s developmental stages of the participants, or the issue of presence.

Gender is another consideration. According to Lenhart, Kahne, Middaugh, Macgill, Evans, and Vitak (2008), 99% of boys and 94% of girls play video games. (p. 2). From a sample of 1102 teens ages 12-17 they report 65% of the daily
gamers were boys, and 35% were girls. What is not clear is whether and how middle school age boys or girls experience or express presence as manifested through game play.

**Design and Implementation of Modules for Assessing Science in Middle School**

SAVE Science is a five-year, National Science Foundation-supported research project in which we are developing a series of VE-based game modules to test whether a VE can be used to assess middle school students’ understanding of scientific content and inquiry. The typical assessment of science knowledge in the United States uses objective type test items tied to a scenario presented in written format with minimal pictures. These types of tests are expected to assess content knowledge, but indeed also test for reading skills. Hence, they are difficult for children who cannot read, are not reading on grade level, or have limited English proficiency as an English language learner (ELL). The SAVE Science project is investigating whether students perform better on situated, context-based assessments as opposed to the traditional text-based tests that test reading skills as much as science.

The SAVE Science assessments are designed to assess individual student’s understanding of local district middle grades (11-14 year olds) curricula. Working with district administrators and teachers, we have identified and are designing around concepts that are currently not well-assessed on state and district high stakes tests. Our design process has two interconnected aspects: the problem scenario that tests individual concepts and the instructional game design which gives context to the problem scenario and provides immersion and presence for the participants. At the time of this study, we have designed two SAVE Science assessment modules for seventh graders based in a medieval world called Scientopolis (see Figure 1). Students work individually through the modules because they are designed as assessments, or tests, where students do not collaborate with other students to solve the problem, just as they would in a test in school.

In both assessment modules, students are met by a local computer-based character who presents the problem scenario to students and creates the motivational prompt for them. In the first module, “Sheep Trouble,” students are asked to
help a farmer find evidence to support a scientific hypothesis for why his new flock of sheep is not doing as well as his original flock. As an additional motivation, students are told that if they fail to find scientific evidence for the problem, the town’s executioner will execute the new sheep to prevent their “bad magic” from spreading to other farms. This module requires students to apply their understanding of beginning speciation and aspects of scientific inquiry to help the farmer with this problem. They can collect evidence from two characters, multiple sheep and environmental clues using various tools (see Figure 2).

The second module, “Weather Trouble,” assesses understanding of weather fronts and different aspects of scientific inquiry. Similar to Sheep Trouble, students are met by a farmer when they enter Scientopolis and asked to help him save his town where a long-lasting drought is sending townspeople away due to fear that the drought is never going to end. Students are asked to investigate the causes of the drought, and to predict if the drought will end soon. Students have a different array of tools to access and a different part of Scientopolis to explore (see Figure 3).

![Figure 3. Weather Trouble overview](image)

Students are given a class period to explore, interact with other computer agents, virtual tools, and objects in the world in order to form a hypothesis of the cause of the problem. Once they have solved the problem, they report back to Farmer Brown.

**Problem definition, Research questions, and contribution**

SAVE Science is a virtual environment designed to assess middle school students’ knowledge and application of science inquiry to a problem set in context. While not designed specifically with “presence” in mind, we hypothesized that the more students perceive themselves as part of the story, the more likely they will actively engage in the activities. This project adds to the literature on virtual games by addressing games for assessment, and by looking at how middle grade students experience and report presence in these environments. Through the implementation of SAVE Science assessment modules in sixth and seventh grade classrooms, we asked the following questions:

- How do middle grade students respond to an online survey about their experience with a SAVE Science module compared to a group discussion of that experience with their teacher?
- Is there a difference between boys and girls in their responses based on prior experience with console or computer games?
- Is there a difference in perceived presence based on age or grade level as demonstrated through an online survey and/or discussion with their teacher, using Piaget’s theory as an interpretive lens?

**Methods**

**Protocol**

This study was a pre-experimental design, as no “experiment” was conducted (Gay, Mills, & Airasian, 2009). Participating students were from two different schools and teachers. Prior to participating in SAVE Science, parents and students submitted signed consent forms to participate in the project, and to be audio and video recorded.
Teachers underwent eight hours of professional development that covered the science topics underlying the modules, the motivation for SAVE Science, plus the logistics and details of implementation. Students were asked to complete a pre-survey that asked them about prior experience with computer and console games, and then completed an introductory module to help them become familiar with the game interface. Following that, teachers implemented one of the two assessment modules, Sheep Trouble (seventh grade class) or Weather Trouble (sixth grade class), with their students. A post-module survey that specifically asked students about their experiences/perceptions of “presence” while completing the module (items were based on Lombard & Ditton, 1997) completed the implementation. Students were asked to participate in a post-module discussion with the teacher that was based on four questions:

- What was the problem you were asked to solve?
- How did you go about solving the problem?
- How was this like being a scientist?
- How was this like or different from taking a test?

These questions were important because the first two indicate what the student thought they were doing in the module, the third indicates what the students believe scientists do, and the fourth indicates how the students interpreted the module as a test or not. In this paper we concentrate on responses to the first two questions. All conversations were recorded and transcribed for accuracy. Content analysis of the transcripts was completed looking for evidence of “being there” language, including the use of the first-person ‘I,’ ‘me,’ or ‘we,’ along with other language that demonstrated the students were engaged, immersed, and interactive in the VE (e.g., ‘he said’ or ‘she said’).

Site and Sample

In spring 2010, the SAVE Science project was implemented in seven schools, three in a large urban school district, and four suburban/rural schools outside the urban school district, all in the United States. Middle school science teachers were invited to participate, and ultimately, seven teachers implemented at least one module with their students. Since this was an assessment, teachers implemented each module on unique timeframes depending on when they taught the topics assessed in the modules. The four teachers in the suburban/rural schools were asked to audio record their post-module discussion. Among these four, only two completed the assignment on time and are the focus of this paper.

From these two teachers, a total of 154 students participated (66 male, 88 female). One teacher included five classes of students in her seventh grade science classroom (46 male, 56 female) after they completed the first module (Sheep Trouble), and the other teacher included two sixth grade classes (20 males, 32 females) after they completed the second module (Weather).

Results

Findings from the pre and post module surveys

We asked students about their prior experience with either computer or console games to give a sense of their prior gaming experience in case this factor impacts their sense of presence in a game as suggested by Lombard and Ditton (1997). From the questions on gaming habits in the pre-module survey, students reported varying levels of prior experience with either computer or console games. Table 2 presents the results related to use of computer games. The results are presented in terms of the numbers and percentage of males or females, by grade level, responding to the question.

The data in Table 2 were analyzed by a two-way ANOVA (gender by grade). This produced a significant main effect for grade ($F_{(1,148)} = 10.48$, $p = .001$, $h^2_p = .020$), a non-significant main effect for gender ($F_{(1,148)} = 10.48$, $p = .097$), and a marginally significant interaction ($F_{(1,148)} = 4.02$, $p = .047$, $h^2_p = .026$). As shown in Table 2, the seventh grade students demonstrate a greater prior use of computer games compared to the sixth graders. Follow-up analyses for the interaction indicated that female students had a significantly higher mean when compared to the male students at the sixth grade ($t = 2.28$, $p = .029$), but that the two groups did not differ in the seventh grade.
Table 2. Survey results for use of computer games

<table>
<thead>
<tr>
<th>Level of Use</th>
<th>Sixth Grade Students (1 F no reply)</th>
<th>Seventh Grade Students (1 M no reply)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (total 20)</td>
<td>Female (total 31)</td>
</tr>
<tr>
<td>Never</td>
<td>4 (20%)</td>
<td>0 (00%)</td>
</tr>
<tr>
<td>Rarely (1 or 2 X per month)</td>
<td>6 (30%)</td>
<td>8 (26%)</td>
</tr>
<tr>
<td>Occasionally (1 or 2 X per week)</td>
<td>6 (30%)</td>
<td>11 (35%)</td>
</tr>
<tr>
<td>Frequently (daily)</td>
<td>4 (20%)</td>
<td>12 (39%)</td>
</tr>
</tbody>
</table>

This information tells us that the majority of these students had some experience with computer games, with only 18% overall having no experience at all. It is important to point out that the term ‘games’ was broadly defined and examples were not requested.

Table 3 presents the results related to use of console games. Again, the results are presented in terms of the numbers and percentage of males or females, by grade level, responding to the question.

Table 3. Survey results for use of console games

<table>
<thead>
<tr>
<th>Level of Use</th>
<th>Sixth Grade Students (1 F no reply)</th>
<th>Seventh Grade Students (1 M no reply)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (total 20)</td>
<td>Female (total 31)</td>
</tr>
<tr>
<td>Never</td>
<td>1 (05%)</td>
<td>5 (16%)</td>
</tr>
<tr>
<td>Rarely (1 or 2 X per month)</td>
<td>2 (10%)</td>
<td>7 (23%)</td>
</tr>
<tr>
<td>Occasionally (1 or 2 X per week)</td>
<td>6 (30%)</td>
<td>13 (42%)</td>
</tr>
<tr>
<td>Frequently (daily)</td>
<td>11 (55%)</td>
<td>6 (19%)</td>
</tr>
</tbody>
</table>

A two-way ANOVA (grade by gender) found a highly significant main effect for gender with a large effect size ($F_{(1,148)}=37.015$, $p = .000$; $h_p^2 = .2$). Neither the main effect for grade nor the interaction were significant. As demonstrated in Table 3, more boys reported playing console games, with fewer boys indicating never using console games (Sixth grade=5%, Seventh grade=2%) compared with girls (Sixth grade=16%, Seventh grade=7%), and more boys reported using them daily (Sixth grade=55%, Seventh grade=49%). Girls did hold their own in playing console games monthly or weekly, but clearly boys play console games more often.

Table 4. Post module survey responses about experiencing presence

<table>
<thead>
<tr>
<th>Level of Use</th>
<th>Sixth Grade Students (1 F no reply)</th>
<th>Seventh Grade Students (1 M no reply)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (total 20)</td>
<td>Female (total 31)</td>
</tr>
<tr>
<td>Sense of ‘being there’</td>
<td>3 (21%)</td>
<td>10 (40%)</td>
</tr>
<tr>
<td>Interacted with others in Scientopolis</td>
<td>2 (14%)</td>
<td>11 (44%)</td>
</tr>
<tr>
<td>Felt people in Scientopolis talked directly to them</td>
<td>4 (29%)</td>
<td>8 (32%)</td>
</tr>
<tr>
<td>Felt they were all together in the game with others in Scientopolis</td>
<td>4 (29%)</td>
<td>8 (32%)</td>
</tr>
</tbody>
</table>

Since responses to the survey items related to presence in the game are not intervaly scaled, a chi-square analysis was used. Answers to the four most important items in the post module survey questions about whether students experienced presence in the game seem to indicate their experiences were mixed, and are presented in Table 4. The first question asked to what extent students had a sense of ‘being there’ inside the Scientopolis game module. Overall, 55% of all the students indicated some level of ‘being there’. A chi-square analysis was used to compare males and females in reporting presence, based on the previous analysis which showed gender to be more significant than grade. In this case, there were no significant differences found ($\chi^2(3)=3.876$, $p=.275$). A second chi-square analysis used to compare sixth and seventh grades found significant differences ($\chi^2(3)= 12.642$, $p = .005$), where seventh graders were more likely to indicate a sense of ‘being there’, while sixth graders were more likely to indicate they ‘felt they were all together in the game with others in Scientopolis’.

59
A correlation analysis was used to determine whether there was a relationship between prior use of either console or computer games and responses to whether students experienced presence in the assessment modules. This analysis used all 10 items specific to presence in the online survey correlated with responses to use of computer games and console games. There were no significant correlations between prior use of console games and students responses to being present in the SAVE Science modules. There were several significant relationships between prior use of computer games and responses to a sense of ‘being there’ inside Scientopolis ($r = .172$, $p = .001$), interacted with others in Scientopolis ($r = .164$, $p = .002$), made eye contact with others in Scientopolis ($r = .172$, $p = .001$), and smiled in response to someone in Scientopolis ($r = .206$, $p = .000$). Students who used computer games prior to a SAVE Science module were more likely to indicate they acted as if they were present in the game environment.

Findings from the post-module discussion

The post discussion session was intended to assist the teachers’ attempts to connect the VE module back to classroom activities and discussions for the students. In this post module discussion, the teacher asked a number of questions about the students’ experience in the game module (see 3.0 above). Analysis of this discussion used the schema presented by Slater and Usoh (1993), where the ‘student talk’ was analyzed for iterations of first (I, we), second (you), or third (he, she, it, they) person references to experiences in the VE. The more first or second person references made, the more likely the students identified with the VE, hence more likely ‘presence’ had been achieved.

While several of the students discussed issues of logistics in moving around in the VE, especially the second class of sixth graders who seemed to fixate on the logistics and game design features, overall students participating in the post-module discussion indicated their actions were akin to “being there” in the VE with Farmer Brown and other characters in both modules. Specifically, in response to the questions about what was the question/problem, and how did they try to solve the problem, students used the words “I,” “you” and “we” often to describe their actions in the module, along with phrases such as “he said” and “she said.” Two-thirds of the total comments included such utterances, with the remaining one-third comments merely stating facts to answer direct questions by the teacher. Students’ using these terms suggest they had identified with the VE enough to self identify with their avatar so they were “talking” or “interacting” with the characters in the VE. Table 5 summarizes the types of “talk” from the students.

<table>
<thead>
<tr>
<th>Category of statements made by students</th>
<th>Number of instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factual (statements of facts relative to the module)</td>
<td>29</td>
</tr>
<tr>
<td>“I” statements</td>
<td>35</td>
</tr>
<tr>
<td>“We” statements</td>
<td>3</td>
</tr>
<tr>
<td>“You” statements</td>
<td>13</td>
</tr>
<tr>
<td>Total number of statements</td>
<td>80</td>
</tr>
</tbody>
</table>

Examples of students “being there” include:

Children who worked with module 1, Sheep Trouble, said things like (seventh graders, ages 12-13):

“I talked with the farmer and the townspeople.”
“I went to like all the sheep and I noticed that all like the new ones were really skinny but then like all the older ones were like really fat.”
“You can measure it’s weight, height, if it’s a male or female.”
“You’re solving someone’s problem.”
“We used the tools to help us.”

Children who worked with the module 2, Weather Module, said things like (sixth graders, ages 11-12):

“I went to the lady with the newspapers and I talked to each other person…”
“I went to each town and used the sci tools and I recorded like the temperature and wind speed.”
“I like went to the newspaper lady. I wrote down like the information on the newspapers. I would write the heading and write down the wind speed.”
“You had to figure out why Scientopolis was having a drought and you had to go to each different place and figure it out.”
“When you’re at Shoretown, there’s a police officer saying that it doesn’t rain most of the day.”
“He asked if was some magic, and I said that no it wasn’t. I said that they were gonna get rain soon.”

Factual statements were related to facts about the modules (e.g., it was raining, the wind was blowing). As noted above, the children who participated in the post module discussion with the teacher gave varying indications of their identification (i.e., achieving presence) with the characters and other objects in the VE. With over two-thirds of all student talk including references to first and second person action, these students demonstrated some level of achieving identification with their avatar and other characters in the VE, which Slater and Usoh (1993) suggest indicates achieving presence in the VE.

**Conclusion**

As noted in the introductory remarks of this paper, prior research literature on presence has been directed toward adults or older secondary age students, and not middle grade students. Given Piaget’s theory (1953) of child development, would we find that children in either the Concrete Operations stage (ages 7-12) or Formal Operations stage (ages 13-adult) would express feeling presence in the SAVE Science modules? Using the narrative analysis scheme put forth by Slater and Usoh (1993), we analyzed the student talk during the post module discussion with the teacher to determine whether this was achieved.

First, it is important to note that there were differences among the children sampled in this study. Prior research indicated males are more likely to use both computer and console games, and this appears to be true for the males in this sample. However, it is also important that the girls were very active in using computer games, and did use console games as well. Further, the children completed two different assessment modules.

In reviewing the survey data, students in grades 6 and 7 were able to take the questions seriously and to answer consistently. In addition, the data analysis showed that for this group of students, prior use of computer games did correlate positively with reported levels of presence in the SAVE Science modules, but not with prior use of console games.

We used the qualitative data to lend further insight into this relationship. These results indicate that presence theory applies to sixth and seventh grade children when using an immersive VE game to assess understanding of scientific inquiry, but in a cautious way. The quantitative data showed that 55% of the students indicated perceiving being in the game world. The sixth graders were more inclined to indicate ‘feeling’ the characters in the game were looking and talking directly to them. The seventh graders were more circumspect and less likely to report the same. In the discussion with the teachers, these students in both grades clearly stated that they (in first and second person) talked with the characters in the game, used the tools afforded to them within the game, took notes in the game, graphed information they collected in the game, and came back to the farmer in both modules with answers to the questions, not always correct, but answers nonetheless.

Why the difference in the data? One potential variable would be past experience with computer/console games, but this appears to not be important since students of both teachers had similar experiences with both. One source may be the survey taken immediately after participating in the game. The second may be the discussion with their teacher immediately after completing the survey. Students may have felt more comfortable saying things to their teacher, were showing off to others in the class, or may not have fully understood the questions in the survey.

The differences in responses to the surveys and the discussion may, or may not, be related to the different stage of Piaget’s development based on age (sixth graders are typically 11-12 years old, while seventh graders are 12-13 years old). The survey responses are more difficult to interpret since there were no significant differences between grades overall. The responses by the sixth graders in the post-module discussion are a bit more concrete, as in responding according to direct perceptions and not separating abstract from real. It was one of the sixth grade classes which could not get past issues about the logistics and game elements, attending more to concrete aspects of playing the game. The seventh graders appear to be operating more abstractly and being able to separate real from abstract. Given the fact that we did not assess these students’ stage of development we cannot say this for sure. However, given their responses to the survey items and the post-module discussion, it seems that the sixth grade students were more concrete in their reactions to the SAVE Science module, not being able to get past the technical aspects of
playing the game. This would mean Piaget’s theory of stages of development, as defined above, fit these students’ responses.

What we can say from this study is that students in sixth and seventh grades are capable of completing an online survey and having a post-module discussion with their teacher about their experiences within an immersive VE module. From these two sets of data (surveys and post-module discussion talk), different impressions of the students’ level of perceived presence were found, which may relate to their level of development as defined by Piaget. From the surveys, the students demonstrated no difference between grades and genders for indicating level of presence given the language they used to express their experiences within the game. They were more likely to use first and second person language to describe their interactions with the characters in the modules than impersonal third person language. The difference in these findings may be due to impersonal nature of the survey where responses were not acknowledged versus the personal discussion with their teacher who responded to their conversation in turn.

Limitations and future research

It is important to point out some limitations to this research. First of all, there was no control group. All students participated by playing one of the games developed through the SAVE Science project. Hence, there is no way to demonstrate that this group of students was different or unique as sixth and seventh grade students. Secondly, they all were from non-urban setting schools. While the demographics of these students is not fully known, the teachers reported verbally that the majority of their students (e.g., >90%) were Caucasian, native English speakers who were not identified as Special Education students. These reported demographics indicate the findings are not generalizable beyond this group of students.

In addition, we did not administer any instrument to determine the exact level of Piaget’s child development for these students. We are surmising that students who are in sixth grade will more likely be in the Concrete Operations stage, and those in seventh grade will be moving into the Formal Operations stage, both based on age. However, it is important to note that without formal assessment, we do not know exactly in which stage of development these children fall.

Acknowledgement

We acknowledge the post-doctoral project manager, Uma Natarajan, along with graduate research assistants—Angela Shelton, Amanda Kirchgessner, Chris Teufel, Tera Kane and Kamala Kandi—for their help in both transcribing these recordings as well as helping to analyze the data. In addition, we thank Dr. Joseph Ducette for his assistance with statistics. Further we thank the teachers who so enthusiastically engaged in implementing SAVE Science assessments with their students, and continue to do so. This material is based upon work supported by the National Science Foundation under Grant No. 0822308.

References


Any Effects of Different Levels of Online User Identity Revelation?

Fu-Yun Yu
Institute of Education, National Cheng Kung University, Tainan City 701, Taiwan // fuyun.ncku@gmail.com

ABSTRACT
This study examined the effects of different levels of identity revelation in relation to aspects most relevant to engaged online learning activities. An online learning system supporting question-generation and peer-assessment was adopted. Three 7th grade classes (N=101) were assigned to three identity revelation modes (real-name, nickname and anonymity) and observed for six weeks. A pretest-posttest quasi-experimental research design was adopted. Findings did not confirm that different levels of identity revelation affected participants' academic performance, nor led participants to view the peer-assessment strategy, the interacting parties, interaction processes, or engaged activities in different ways. Implications for generalizability of research findings and suggestions for teaching practices are offered.

Keywords
Anonymity, Computer mediated communication, Nickname, Peer assessment, Question generation

Introduction
According to the information processing theory, student question-generation may help students to reflect on received information, and elaborate and transform that information into inter-connected forms. Additionally, from the perspectives of constructivism and metacognition, question-generation may induce students into a habitual state of constructing personally meaningful knowledge and employing various metacognitive strategies (Yu, Liu, & Chan, 2005). Empirically, the cognitive, affective and social benefits of student question-generation, such as the enhancement of learners' comprehension, cognitive strategies, metacognitive strategies, creative thinking, interest, confidence and communicative process within groups have been asserted by numerous researchers (Abramovich & Cho, 2006; Barlow & Cates, 2006; Brown & Walter, 2005; Fellenz, 2004; Whiten, 2004; Wilson, 2004; Yu, 2005; Yu & Hung, 2006; Yu & Liu, 2005). Additionally, by revealing insights into students' abilities in the subject content and providing an accurate assessment of what learners are capable of accomplishing, student question-generation holds benefits for the implementing teachers for its assessment value (Whiten, 2004).

While the theoretical and empirical bases support the student question-generation approach to teaching and learning, the fact that most students have not experienced it during their formal course of study (Moses, Bjork, & Goldenberg, 1993; Vreman-de Olde & de Jong, 2004), and that students being introduced to student question-generation activities expressed concerns regarding their ability to construct good questions, (Yu & Liu 2005) warrant serious attention. To assist its adoption and diffusion in classrooms, the development of an online question-generation learning space with several types and levels of support (i.e., peer-assessment and different identity revelation modes) has been the focus of one research project since 2006.

Why student question-generation in online space?
As computer and telecommunication technologies converge, web-based learning systems present themselves as promising learning tools for the 21st century. There are several advantages to online learning activities beyond the attributes that are frequently associated with computers (e.g., immense storage space, high processing speed, multimedia appeal and capability, time- and space-independence). These include: the potential for socially active learning, ease of information management from dispersed locations, flexible functionality, and customizable data (Smaldino, Lowther, & Russell, 2007; Newby, et al., 2006).

It is evident that student question-generation activities implemented online offer several distinctive advantages. Explicitly, construction of knowledge in the form of questions and answers and associated activities (e.g., discussion, sharing and modeling) can be effectively and efficiently carried out in network-mediated space. Furthermore, all artifacts produced during the process can be kept in a “learner portfolio” and made accessible easily for future reference (e.g., different versions of generated questions, interactions between authors and assessors about generated questions). Moreover, students can incorporate multimedia files (including graphics, animation, sound and video clips) as part of their generated questions or feedback. Additionally, personalized automatic notifications can be used...
to alert interacting parties of updated messages (in this case, newly posed responses to questions or assessments). Finally, system functions can be dynamically changed to either suit the instructors’ educational goals and instructional plans, or match students’ developmental stages, needs and preferences (Yu, 2009). All in all, these guarantee more fluent question-generation for implementing teachers and students alike.

Why peer-assessment in conjunction with student question-generation?

Several cognitive processes are mobilized when students engage in peer assessment activities. Assessing peers’ work stimulates critical thinking as constructive comments and objective judgments are the targeted outcomes. Reviewing peers’ work in turn frequently re-directs students to re-examine their own work and the follow-up enhancements or modifications. On the other hand, when students receive feedback from assessors, provided comments may introduce cognitive conflict and challenge students to deal with incomplete or inaccurate conceptualizations. Knowledge structuring and re-structuring are cultivated through a continuous process of self-examination, monitoring, evaluation, correcting, adjustment, among other things. These processes, based on cognitive conflict theory, social constructivism and social learning theory, should promote cognitive abilities and critical thinking (Falchikov & Goldfinch, 2000; Topping, 1998; van Gennip, Segers, & Tillema, 2009). Empirical evidence further supports peer-assessment’s facilitative effects on promoting learners’ higher-order thinking, cognitive re-structuring, motivation, academic performance and attitudes toward studied subject (Brindley & Scoffield, 1998; Falchikov & Goldfinch, 2000; Gatfield, 1999; Hanrahan & Isaacs, 2001; Liu et al., 2001; Purchase, 2000; Topping, 1998; Tsai, Lin, & Yuan, 2002; van Gennip et al., 2009).

The practicality of peer-assessment also adds value to question-generation activities. Providing feedback to individual students about their work (in this case, questions generated by students) is important, but it is very time-consuming and effort-intensive for teachers when it is solely their responsibility. In view of this, making use of peers as assessors allows for timely and personalized feedback while allowing instructors more time to focus on other aspects of the class.

Why different online identity revelation modes?

Different levels of identity revelation can be made available online: real-identity (complete identity revelation), created identity via nickname (partial identity revelation) and anonymity (complete identity concealment). According to social psychology literature, evaluation anxiety and self-validation based on social comparison may be of less concern to individuals participating in situations where they are not identified (Cooper et al., 1998; Franzoi, 2006; Lea, Spears, & de Groot, 2001; Moral-Toranzo, Canto-Ortiz, & Gómez-Jacinto, 2007; Palme & Berglund, 2002; Pinsonneault & Heppel, 1997-98; Postmes & Lea, 2000). By lessening inhibitions, anonymity has been suggested to permit group members to meet needs that they cannot otherwise satisfy, and promote intimacy, affection, and playfulness (Festinger, Pepitone, & Newcomb, 1952; Gergen, Gergen, & Barton, 1973).

Nicknames, on the other hand, have been reported to possess great motivational potential. Like anonymity, nicknames can protect participants from being identified immediately. The flexibility, easiness and fun of changing one’s identity to suit his or her mood at any given time holds further motivational value for participants (Yu & Liu, 2009). In view of its popularity among web-users and its prevalence in newsgroups, online chat-rooms, forums and instant messaging space, the potential of identity self-creation, as compared to real-identity and no-identity (anonymity), demands rigid investigation to warrant its inclusion and use in educational contexts.

Purpose of this study

While most research found anonymity to be statistically significantly different from identified situations with regards to perceptual impression, communication and behavior (Cooper et al., 1998; Moral-Toranzo et al., 2007; Pinsonneault & Heppel, 1997-98; Postmes & Lea, 2000; Yu, 2003; Yu, Han, & Chan, 2008), and students exhibited significantly varied preferences for different identity revelation modes (Yu & Liu, 2009), it has not yet been found whether different identity modes have different educative effects.
Considering that identity concealment, instant creation and re-creation is one of the prominent features afforded by networked technologies, the effects of three identity revelation modes—real-name (real identity), nickname (created identity of the users’ choice) and anonymity (concealed identity), on aspects most relevant to the engaged activity were examined: academic achievement, attitudes toward the peer-assessment strategy and perceptions toward interacting parties, interaction process and the question-generation and peer-assessment activity.

**Methods**

**Participants and context**

A total of 101 seventh graders from three classes taught by the same instructor in southern Taiwan participated in the study for six consecutive weeks. The study took place in a “Science and Technology” course. The study started right after the first exam and ended prior to the second exam at the second semester of the school year.

For the duration of the study the students had class in a computer laboratory once a week where they participated in an online question-generation and peer assessment activity after attending three 50-minute instructional sessions allocated for biology. Four chapters on the laws of genetics, human inheritance, biotechnology, genetic consultation and evolution were covered during the study.

**Online question-generation learning system**

A learning system called the Question Authoring and Reasoning Knowledge System (QuARKS) that allows students to contribute to and benefit from the process of question generation and peer feedback was adopted for use in the study. Essentially, QuARKS is comprised of two sub-systems—question authoring and question reasoning. For a detailed description on the system, refer to Yu (2009).

**Question authoring**

In QuARKS students can author several types of questions including true-false, matching, fill-in-the-blank, multiple-choice, short-answer, and so on. Students need to fill out several fields to complete a successful submission. For instance, for multiple-choice question-generation, students need to provide a question-stem, four alternatives, an answer key, and annotation for each question posed. All questions contributed by students are kept in an item bank database, waiting to be evaluated by peers and re-defined by the author of the question in the follow-up question reasoning phase.

**Question reasoning**

QuARKS employs a question reasoning system to enhance interaction, collaboration and negotiation of meaning between question-authors and their peers (assessors). Assessors give evaluative feedback using an assessor-to-author assessment form. Assessment criteria associated with different types of questions are provided through a pull-down menu to foster focused, objective and constructive communication. Once feedback is received, question authors can respond to them via an author-to-assessor form.

**Experimental design and conditions**

A pretest-posttest quasi-experimental research design was adopted and three treatment conditions with different levels of identity revelation were devised for the study. The levels of identity concealment ranged from complete concealment (i.e., the anonymity group) to partial concealment (i.e., the nickname group where the degree of disclosure depends on how much personal information each individual is willing to reveal) to full disclosure (i.e., the real-name group).

In the real-name condition, the student’s full name was automatically retrieved from the database and shown at the top of the field where questions and feedback were viewed by both assessors and authors, respectively (See Fig. 1).
Figure 1. Real-name mode

Figure 2. Nickname mode

Figure 3. Anonymity mode
In the nickname group, the student’s created identity was shown at the top of the field containing his or her generated questions and comments (See Fig. 2). Students are free to change their nicknames to reflect their current state of mind each time they construct or assess a new question. In the anonymity group, information on the question-author or assessor was not shown, and only the word “anonymous” appeared at the top of the question and comment field. (See Fig. 3).

Experimental procedures

Three intact classes were randomly assigned to different treatment conditions. Considering that true/false, fill-in-the-blank and multiple-choice questions are among the most frequently encountered question types in middle schools, they were adopted for the study.

To ensure that participants possessed the fundamental skills needed for the generation and assessment of these types of questions, a training session with hands-on activities was held at the beginning of the study, in addition to a lesson on the operational procedures of QuARKS. One pamphlet containing (a) learning objectives, (b) QuARKS key features and functions, (c) question generation criteria and sample questions, and (d) peer assessment criteria and sample feedback/comments, was distributed for individual reference.

During each weekly online learning activity, students were first directed by the instructor to individually compose at least one question of each type in accordance with the covered instructional contents. Each student then individually assessed at least one question from the pool of peer-generated questions for each question type.

To establish a baseline regarding students’ perceptions of the different aspects of the activity, a real-identity mode was used for all conditions during the first two sessions. Afterwards, students in different conditions used their respective identity revelation modes. Students’ performance on the first biology exam was collected. A questionnaire about the examined variables was disseminated for individual completion before different treatment conditions were implemented in different groups, which started at the third week. After exposure to the activity for six weeks, students completed the same questionnaire. Students’ performance on the second biology exam was then also collected.

Measurements

The effects of different identity revelation modes on students’ academic performance were assessed by the first and second biology exams of the participating school. Generally, the three exams spaced evenly throughout a 5-month long semester (i.e., approximately six weeks separation between two exams) are arranged by schools and administered at the same time for all students and all major subjects at high school levels in Taiwan. Item analyses were conducted which ascertained that the test items correctly discriminated between high and low scores (average 0.61 and 0.44 for the first and second exam, respectively) and that the items as a whole were of moderate difficulty (average 0.66 and 0.63 for the first and second exam, respectively).

The effects of different treatment conditions on students’ perceptions of various aspects of the engaged activity were assessed using the same pre- and post-questionnaire, which consists of a set of four 5-point Likert scales (5=strongly agree, 4=agree, 3=no opinion, 2=disagree, 1=strongly disagree). Existing instruments on related areas (i.e., peer-assessment, perceptions toward interacting parties, perceptions toward communication process and perceived learning environment) were referred to, and items were adapted to fit the targeted experimental context. Specifically, for “Attitudes toward Peer-Assessment Scale,” “Peer Assessment Questionnaire” (Brindley & Scoffield, 1998), “Peer Assessment Questionnaire” (Wen & Tsai, 2006) and “Fairness of Grading Scale” (Ghaith, 2003) were referred to. When constructing “Perception Toward Assessors Scale,” “Peer Assessment Rating Form” (Lurie, Noziger, Meldrum, Mooney & Epstein, 2006), ”The Questionnaire on Teacher Interaction” (Wubbels & Brekelmans, 2005), “Student Perceptions of their Own Dyad” (Yu, 2001), “Student Perceptions of Other Dyads” (Yu, 2001) and “Cooperative Learning Scale” (Ghaith, 2003) were used as a reference whereas “Student Evaluations of their Experience of Assessment” (Stanier, 1997), “Student Perceptions of the Communication Process within the Dyad” (Yu, 2001) “Student Perceptions of the Communication Process among the Dyads” (Yu, 2001), “Cooperative Learning Scale” (Ghaith, 2003) and “Peer Assessment Rating Form” (Lurie, et al., 2006) were referred to when it comes to the construction of “Perception toward the Interaction Process with Assessors Scale.” Finally, “Learning
Environment Dimensions in the CCWLEI Questionnaire” (Frailich, Kesner & Hofstein, 2007) and “Clinical Learning Environment Scale” (Dunn & Hansford, 1997) were referred to for “Perception toward the Engaged Activity Scale.”

Before the questionnaire was used in the actual study, a separate group of 242 seventh graders from four different middle schools were recruited to ensure instrument validity and reliability. Only items that passed item analysis, factor analysis and internal consistency were included. Psychometrics data for each of the scales (number of items, total variance explained by the extracted factors and reliability) are listed in Table 1. A complete list of items of the adopted questionnaire was included in Appendices A-D to allow the reusability of the validated measures.

Table 1. The number of items included, total variance explained by factors and the reliability of each of the adopted scales

<table>
<thead>
<tr>
<th>Scales</th>
<th>Number of items included</th>
<th>Total variance explained</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitudes toward peer-assessment</td>
<td>19</td>
<td>56.65%</td>
<td>0.87</td>
</tr>
<tr>
<td>Perception toward assessors</td>
<td>23</td>
<td>64.51%</td>
<td>0.93</td>
</tr>
<tr>
<td>Perception toward the interaction process with assessors</td>
<td>23</td>
<td>64.07%</td>
<td>0.92</td>
</tr>
<tr>
<td>Perception toward the engaged activity</td>
<td>7</td>
<td>68.66%</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Data analysis

Data was analyzed using the analysis of covariance technique (ANCOVA) which increased statistical power by accounting for possible variability caused by pre-existing states associated with each of the observed variables before the inception of different treatment conditions. Explicitly, students’ scores on the first biology exam were used as covariate when analyzing data on student academic achievement, whereas students’ scores on each of the scales of the pre-session questionnaire (i.e., attitudes toward peer-assessment, perceptions toward assessors, perceptions toward the interaction process with assessors and perceptions toward the question-generation and peer-assessment learning activity) was used as covariates for each of the respective data analysis. The test of homogeneity of the within-class regression coefficient was conducted first to ensure that the assumption of parallel of within-class regression slopes for ANCOVA was met. A .05 level of significance was adopted for use in this study.

Results

The test of homogeneity of within-class regression was met for all examined variables: academic performance (F=1.19, p>.05), attitudes toward peer-assessment (F=1.22, p>.05), perceptions toward assessors (F=2.55, p>.05), perceptions toward the interaction process with assessors (F=0.88, p>.05) and perceptions toward the question-generation and peer-assessment learning activity (F=0.35, p>.05).

ANCOVA found that students using the real-name, nickname and anonymity identity revelation modes did not score statistically differently in any of the observed variables: academic performance (F=1.19, p>.05), attitudes toward peer-assessment (F=1.22, p>.05), perceptions toward assessors (F=2.55, p>.05), perceptions toward the interaction process with assessors (F=0.88, p>.05) and perceptions toward the question-generation and peer-assessment learning activity (F=0.35, p>.05).

Table 2. Descriptive statistics of observed variables of three identity revelation modes

<table>
<thead>
<tr>
<th>Observed variables</th>
<th>Treatment groups</th>
<th>Real-Name N=35</th>
<th>Nickname N=34</th>
<th>Anonymity N=32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre M (SD)</td>
<td>59.54 (19.43)</td>
<td>51.77 (20.29)</td>
<td>62.25 (25.52)</td>
</tr>
<tr>
<td></td>
<td>Post M (SD)</td>
<td>61.26 (17.56)</td>
<td>55.53 (19.84)</td>
<td>60.38 (22.39)</td>
</tr>
<tr>
<td></td>
<td>Adjusted M</td>
<td>59.85</td>
<td>60.34</td>
<td>56.81</td>
</tr>
<tr>
<td>Academic performance</td>
<td>Pre M (SD)</td>
<td>68.22 (8.29)</td>
<td>66.68 (12.94)</td>
<td>62.94 (12.44)</td>
</tr>
<tr>
<td></td>
<td>Post M (SD)</td>
<td>71.44 (10.05)</td>
<td>66.74 (12.77)</td>
<td>66.29 (12.85)</td>
</tr>
<tr>
<td></td>
<td>Adjusted M</td>
<td>70.43</td>
<td>66.42</td>
<td>67.62</td>
</tr>
</tbody>
</table>
Discussion & Conclusions

When compared to real identity situations, anonymity has been shown to reduce participants’ unsettled emotional feelings and restraint in their interactions with others (Cooper et al., 1998; Moral-Toranzo et al., 2007; Pinsoneault & Heppel, 1997-98; Postmes & Lea, 2000; Yu, 2003; Yu et al., 2008), or lead to deviant, harmful, and socially undesirable behaviors due to the loss of a sense of self-awareness and individual accountability (DeSanctis & Gallupe, 1987; Kiesler, Siegel, & McGuire, 1984). Nicknames, on the other hand, have been suggested to hold additional motivational value for participants by allowing them to be identified by any codes or symbols of their choice at that point in time (Yu & Liu, 2009). Prior research has found that students exhibited statistically different preferences for the three distinctly different identity revelation modes (with the majority preferred nickname modes most when authoring or assessing questions) (Yu & Liu, 2009). This study was undertaken to further examine if any comparative educative differences exist among these three modes.

The current study did not confirm the researcher’s hypothesis that different levels of identity revelation would affect participants’ academic performance, or any aspects closely related to the engaged activities (including participants’ views toward the peer-assessment strategy, the interacting partners, interaction process, or engaged activity). Despite this, important implications were derived by combing through related literature and closely examining the context of investigation in search for insight and explanations.

A close analysis revealed several differences in structural features between existing studies and the current study. First, existing studies on anonymity were mostly conducted within the framework of computer mediated group work (e.g., electronic brainstorming, decision-making), or competitive gaming environments, whereas the current task under investigation stressed the “mutual helping” (versus competitive) aspect of interaction and for the support of “learning” (versus group work). Specifically, interacting parties in the study were expected to serve as partners in learning and communicate with the intent of providing constructive feedback for the enhancement of questions their peers generated as well as their overall learning associated with the studied contents.

Another dimension separating the present study from existing studies was the constituents of the formed groups. Existing studies mostly involved temporarily formed groups of people whose relationships have not been formed before the intervention while this study examined intact groups of classmates who have known each other for at least one semester at the time of the study.

The third main difference was the duration of the experiment. Existing studies on anonymity were typically conducted over a short period of time (with many CSCW studies adopted a one-shot approach) whereas this study observed students interacting online for six periods, extended over six weeks.

The fourth difference might derive from different cultural settings within which the studies were carried out. The present study was conducted in secondary school settings in an oriental country where collectivism is part of the cultural norm, whereas most existing studies were conducted in western countries where individualism is valued.

As aforesaid, although no differences were observed in any of the examined variables, this study’s contradictory findings to previous studies on anonymity have important implications. Foremost, the results and insights yielded from studies done on groups with primarily oppositional or zero-interdependence relationships among users (usually the case in competitive activities or computer-supported collaborative work, CSCW) who convened temporarily and interacted for a short time, will be inappropriately generalizable to contexts where continuous, reciprocal assistance...
of acquainted partners is valued, as was witnessed in this study.

Finally, though a prior study found that participants revealed significantly different preferences toward different identity revelation modes (Yu & Liu, 2009), in light of the non-significance results of the present study, learning technologists should not feel their need to base their decision on learning systems on the availability of identity revelation mode. Furthermore, instructors should not feel compelled to resort to anonymity or nickname as the mode of choice for the enhancement of learners’ perceptions of the interacting activities or academic performance.

Acknowledgements

This paper was funded by a 3-year research grant from the National Science Council, Taiwan, ROC (NSC 96-2520-S-006-002-MY3). The author would like to thank the instructor, Shannon Huang and research assistants, Meiju Chen, Knem Chen and Yu-Zen Liao for their assistance during the process.

References


Yu, F. Y., Han, C. L., & Chan, T. W. (2008). Experimental comparisons of face-to-face and anonymous real-time team


Appendix A: Attitudes toward Peer-Assessment

1. Peer-assessment assisted my learning.
2. Peer-assessment allowed me to be aware of the main ideas of the studied materials.
3. Peer-assessment enabled me to recognize more of the demands the instructor had for this course.
4. Peer-assessment improved my written communication skills.
5. Peer-assessment allowed me to understand better other classmates’ thoughts.
6. Peer-assessment put me under quite a lot of pressure.
7. Peer-assessment increased my motivation toward learning
8. Peer-assessment led me to like this course better.
9. Peer-assessment enabled me to interact more with the instructor.
10. Peer-assessment allowed me to have a sense of participation.
11. Peer-assessment enabled me to interact more with peers.
12. Assessment should not be part of student responsibility.
13. Instructors should set up explicit rules by which students abide for peer-assessment.
14. Students should involve in the set-up of the rules for peer-assessment
15. Peer-assessment is an objective assessment method.
16. The comments or ratings that peers provided me were fair.
17. Peer-assessment took up much of my time.
18. When giving comments or ratings to peers performance or work, I felt that it was affected by those given to me in the first place by my peers.
19. If the received comments or ratings were lower than what I expected, I would give lower comments or ratings to peers’ performance or work in return.
Appendix B: Perception toward Assessors

1. I did not like my assessors.
2. I had negative feelings towards my assessors.
3. I felt that that my assessors were friendly.
4. I felt that that my assessors were responsible.
5. I felt that that my assessors showed respect, compassion and empathy toward others.
6. I felt that that my assessors did not have much patience.
7. I felt that that my assessors were strict.
8. My assessors led me to feel frustrated.
9. My assessors brought me a sense of pressure.
10. My assessors made me feel afraid.
11. I felt that that my assessors did not like me.
12. I felt that that my assessors were hostile toward me.
13. Most of the comments provided by my assessors were phrased in a negative and critical way, which did not help my learning.
14. I felt that that my assessors liked to help me learn.
15. I felt that that my assessors could help my learning.
16. I felt that that my assessors wanted me to do my best schoolwork.
17. I felt that that my assessors cared about how much I learned.
18. I felt that that my assessors really cared about me.
19. I felt that that my assessors cared about my feeling.
20. I felt that that my assessors liked me as much as they liked others.
21. My assessors displayed insensitivity and lack of understanding for others’ view.
22. My assessors did not have enough knowledge on the process and goal of the engaged peer-assessment activity.
23. I felt that that my assessors did not know enough on the learned subject—biology.
Appendix C: Perception toward the Interaction Process with Assessors

1. I learned from my experiences interacting with my assessors.
2. My interaction with most assessors was enjoyable.
3. The interaction process between my assessors and me was efficient and effective.
4. When a disagreement happened, my assessors and I negotiated for possible ways for further revision.
5. When receiving comments/feedback from my assessors, I was willing to respond to them.
6. When my assessors did not understand the purpose of my constructed question, I would elaborate further.
7. While interacting with assessors, I was willing to share my thoughts in writing with them.
8. I find it hard to state my thoughts clearly when interacting with my assessors.
9. I rarely shared my thoughts and reasoning with my assessors during peer-assessment.
10. I often got discouraged when interacting with my assessors.
11. I often felt upset when interacting with my assessors.
12. When interacting with my assessors, we had quite a few arguments.
13. It was hard to reach a consensus with my assessors.
14. When I could not understand the comments or suggestions provided by my assessor, I was afraid to ask further.
15. Even when I could not agree with the ratings or comments given by my assessors, I was afraid to ask further.
16. My assessors were open and acceptable to my responses to their comments/suggestions.
17. When I did not understand the reasoning behind the rendered comments/suggestions, my assessors would elaborate more when being questioned.
18. My assessors valued my opinions.
19. The information and assistance provided by my assessors were helpful.
20. My assessors did not share information or resources.
21. My assessors lacked of understanding for others’ views.
22. The comments or ratings provided by my assessors on my generated questions were too soft.
23. My assessors were unable to explain their reasoning with regard to provided comments/ratings.
Appendix D: Perception toward the Engaged Activity

1. I like to participate in the online question-generation and peer-assessment activity.
2. Online question-generation and peer-assessment was a good way for my learning.
3. Participating in the online question-generation and peer-assessment activity for me for the most part was very interesting.
4. Online question-generation and peer-assessment activity enhanced my interest to learn biology.
5. This experience I had with the online question-generation and peer-assessment made me more eager to learn biology better.
6. The online question-generation and peer-assessment activity added variety to learning biology.
7. The online question-generation and peer-assessment activity helped students to understand the biology-related topics.
Relationship between Students’ Emotional Intelligence, Social Bond, and Interactions in Online Learning

Heeyoung Han and Scott D. Johnson
Department of Medical Education, Southern Illinois University School of Medicine, Springfield, IL, USA // 1College of Education, University of Illinois at Urbana-Champaign, IL, USA // hhan@siuemed.edu // sjohnson@illinois.edu

ABSTRACT
The purpose of the study was to investigate the relationship between students’ emotional intelligence, social bond, and their interactions in an online learning environment. The research setting in this study was a 100% online master’s degree program within a university located in the Midwest of the United States. Eighty-four students participated in the study. Using canonical correlation analysis, statistically significant relationships were found between students’ emotional intelligence, social bond, and the interactions that occurred naturally in the educational setting. The results showed that students’ ability to perceive emotion by facial expression was negatively related to the number of text and audio messages sent during synchronous interaction. Additionally, the ability of students to perceive emotion was positively related to peer bonding. Lastly, students’ bond to their online program was associated with management type interaction during synchronous discussion sessions. Several implications for online learning practitioners and researchers are discussed.

Keyword
Emotional intelligence, Social bond, Online interactions

Introduction
Interaction is a critical factor in the quality of online learning (Berge, 1997; Fredericksen et al., 2000; Garrison, Anderson, & Archer, 2001; Marks et al., 2005; Swan, 2001; Vrasidas & McIsaac, 1999). Interaction in online learning environments has been found to have a close positive relationship with students’ higher order thinking (Garrison, Anderson, & Archer, 2001) and cognitive learning outcomes (Berge, 1997). Interaction between people, defined as dialogue, facilitates deep and reflective learning for the purpose of achieving learning goals in social learning environments (Berge, 2002; Mayers, 2006), which functions as a decisive factor in decreasing transactional distance (Moore, 1997).

Given that emotion, cognition, and behavior are highly interdependent (Cornelius, 1996; Planalp & Fitness, 1999), students’ interaction can be understood in an emotional dimension as well as a cognitive dimension. Emotion has received attention as a critical element of social interaction in the communication field (Andersen & Guerrero, 1998; Burleson & Planalp, 2000; Planalp & Fitness, 1999). In the field of education, emotions have been found to affect students’ cognitive learning as well as teachers’ instructional behavior (Pekrun et al., 2007). Consequently, emotional intelligence has been discussed as one of the important intelligences and competencies to promote and regulate personal intellectual growth and social relational growth (Mayer & Salovery, 1997). While the definitions and constructs of emotional intelligence are varied, emotional intelligence is defined as a set of abilities which involves operating emotional information that represent emotional signals (Mayer et al., 2004). If emotional intelligence is a critical competency for understanding student learning experiences, then students’ emotional intelligence might be one of the areas to be investigated to better understand students’ online learning experiences.

The emotional dimension of interaction should be explored along with the social dimension. Given that positive and constructive interactions can be achieved by respectful and active participants (Moore, 1997), students’ emotional and social relationships are believed to promote interaction (Holmberg, 1991). While many scholars have stressed the social dimension of interaction (Berge, 1997, 2002; Garrison, Anderson, & Archer, 2001; Holmberg, 1991; Lave & Wenger, 1991; Moore, 1983, 1997; Wagner, 1994), there have been few attempts to extend our understanding of interactions in online learning along emotional and social dimensions.

Social bonding theory can be applied to understanding the emotional and social dimension of students’ interactions. Social bond theory was initially proposed to understand an individual’s antisocial behaviors such as delinquency or crime in sociology (Hirschi, 1969). Later, social bond theory was applied to explain social and emotional learning (Newmann et al., 1992; Wehlage et al., 1989; Zins et al., 2004) in K-12 school environments. Student participation and engagement in school activities are used to represent social and emotional learning (Wehlage et al., 1989). Some
studies have found positive relationships between student social bonding and school effectiveness including academic achievement and school engagement (Leibold, 2000; Newmann et al., 1992; Pryor, 1994; Wehlage et al., 1989). From this theoretical perspective, positive interactions between instructors and peers can reinforce their emotional and social bonding as well as their attachment to their online learning program, which leads them to be motivated to accept and implement norms and values of social agents (Catalano & Hawkins, 1996; Hirschi, 1969; Wehlage et al., 1989).

Based on the literature, this study proposed a conceptual framework to represent emotional and social learning for understanding online interactions (see Figure 1). There are three dimensions in the conceptual framework. One is students’ emotional ability to perceive, use, understand, and manage emotions, which represents an individual’s emotional intelligence. A second dimension is the degree to which students are emotionally and socially attached to their online program, their instructor, and their peers, which represents their social-psychological attachment. The third dimension is the online interactions that student have, which represents their cognitive and behavioral involvement in an online learning environment.

![Figure 1. Conceptual Framework for Emotional Intelligence, Social Bond, and Interactions in Online Learning](image)

**Purpose of the study**

The purpose of this study was to investigate the relationship between students’ emotional intelligence and their interactions in both synchronous and asynchronous online learning environments. The main focus of the investigation was the extent of the relationship between the three dimensions of emotional intelligence, social bond, and interaction.

**Research questions**

In order to investigate the problem, the following research questions were addressed.

- What is the relationship between students’ emotional intelligence and their degree of social bond in online learning environments? What is the most important variable in the relationship?
- What is the relationship between students’ emotional intelligence and interactions in online learning environments? What is the most important variable in the relationship?
- What is the relationship between students’ degree of social bond and interactions in online learning environments? What is the most important variable in the relationship?

**Method**

The study used an ex post facto design and correlational analysis to discover the statistical relationships between students’ emotional intelligence and the interactions that occurred naturally in an online learning environment.
Research participants

The target population of this study was graduate students enrolled in an online master’s degree program in a university located in the Midwest region of the United States. Eighty-one students out of a total of 188 enrolled students agreed to participate in the study. The students’ online learning system utilized Moodle for the asynchronous interactions and Elluminate for the synchronous interactions.

Data sources

In order to measure emotional intelligence, the Mayer-Salovey-Caruso Emotional Intelligence Test (Mayer, Salovey & Caruso, 2001; MSCEIT V2.0) was administered to the 81 participants. The test contains 141 items and consists of four branches, which include Perceiving Emotion (EI-B1), Using Emotion (EI-B2), Understanding Emotion (EI-B3), and Managing Emotion (EI-B4). Each branch uses two different tasks to measure each construct: Perceiving Emotion [Face task (EI-A) and Pictures task (EI-E)], Using Emotion [Facilitation task (EI-B) and Sensation task (EI-F)], Understanding Emotion [Changes task (EI-C) and Blends task (EI-G)], Managing Emotion [Emotional Management task (EI-D) and Social Management task (EI-H)]. The MSCEIT has high construct validity and moderate content validity, predictive validity, and external validity (McEnrue & Groves, 2006). Additionally, MSCEIT has been found to have high reliability (Lopes et al., 2003).

In order to measure social bond, the Social Bonding Scales (SBS) from the Wisconsin Youth Survey (Wehlage et al., 1989) was administered with a slight revision for the adult participants of online courses. Factor analysis was conducted in order to identify common factors that underlie the social bond variable. The factor analysis showed that nine items did not represent the three distinct social bond factors. After the nine items were removed from the scales, the remaining 16 items were found to produce three very distinct factors in terms of bond to peer, bond to instructor, and bond to program. These 16 items were then used to measure social bond to peer, instructor, and program for the current study. The 16 items of the revised Social Bond Scale had a high reliability for peer ($\alpha$=.680), instructor ($\alpha$=.885), and program ($\alpha$=.822), which are higher than the 0.60 required for Cronbach’s $\alpha$.

For this study, the interaction variable is operationalized as students’ amount of messages and types of messages in their synchronous and asynchronous communication (Bannan-Ritland, 2002; Kearsley, 2000). Text and audio interaction data were collected from ten synchronous and asynchronous sessions archived in Elluminate and Moodle. The amount of interaction was determined by the number of text messages (MT), the number of audio messages (MA), the length of the text messages (WT), and the length of the audio messages (WA). The number of words was counted for the length of the text messages (WT) and the length of the audio messages (WA). As for types of interaction data, content analysis of messages was conducted to classify the types of messages that students posted. Students’ messages were coded into three types: Work, social, and management (Yoon, 2006). Work type interaction contains messages for goal-directed activities as students complete tasks and elaborate their ideas in discussions. Social type interaction includes messages to share their personal experience and to build social relationships with other people. Management type interaction includes messages for social bonding, addressing workload, and reporting problems in order to complete tasks. In order to increase reliability of the coding, two individuals were hired for the data coding. Consistent agreement among coders was 63% during the initial coding and increased to 94.8% through comparison and discussion.

Data analysis

Canonical correlational analysis was conducted using SAS 9.1 to identify the relationship between the emotional intelligence variables, the social bond variables, and the interaction variables. Because Canonical correlational analysis examines the relationship between pairs of variables (Rencher, 2002), three canonical correlation analyses were performed. Emotional intelligence, social bonding, and interaction are considered latent variables, each with a set of measured indicator variables. In emotional intelligence, four branch scores, including perceiving emotion, using emotion, understanding emotion, and managing emotion, became the indicator variables. Additionally, eight task score variables under the four branch scores were also considered as indicator variables. For social bonding, bonding to peers, bonding to online program, and bonding to instructor became the indicator variables. For interaction, the number of messages and the number of words were the indicator variables for the amount of interaction. For the types of interaction, the number of work type of interaction, the number of social type of
interaction, and the number of management type of interaction were the indicator variables. The age variable was controlled in the associations of interaction with emotional intelligence. The gender variable was controlled in the associations of emotional intelligence with social bond and interaction.

Results

The participant ages ranged from 24 to 63 (M=40.5 SD=10). The majority of the participants were female (74%), native English speakers (92%), and Caucasian (87%). The participants represent the population in the program because the population was also primarily female (67.5%), Caucasian (74.8%), and their ages ranged from 22 to 63 (M=38.33 SD=9.12).

Relationship between emotional intelligence and interaction

The results showed a negative relationship between emotional intelligence and the amount of interaction in synchronous interaction. Among the variables, students’ ability to perceive emotion by facial expression and the number of text and audio messages in synchronous interaction were found to be contributing variables in the negative association. No relationship was found in asynchronous interactions. The results showed no significant canonical correlation between students’ emotional intelligence and types of interactions both in synchronous and asynchronous online learning environment.

After examining the Pearson partial correlation, canonical correlation analysis was conducted to further examine the relationship between the two sets of research variables. Two indicator variables in the managing emotion branch were excluded in the canonical correlation analysis because they do not have Pearson partial correlation with any interaction variables, which may weaken the canonical correlation between the latent variables. As shown in Table 1, the first canonical correlation between emotional intelligence and the amount of messages in synchronous interaction was found to be statistically significant (r=.506, p=.048, α =.05). The squared canonical correlation for the first dimension was .256, which represents the amount of shared variance between the amount of interaction canonical variable and the emotional intelligence canonical variable.

<table>
<thead>
<tr>
<th>#</th>
<th>Canonical Correlation</th>
<th>Squared Canonical (R²)</th>
<th>Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)</th>
<th>Test of H0</th>
<th>Eigenvalue</th>
<th>Cumulative</th>
<th>F</th>
<th>DF 1</th>
<th>DF 2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.506</td>
<td>.256</td>
<td>.335</td>
<td>.602</td>
<td>1.54</td>
<td>24</td>
<td>238.43</td>
<td>.048*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.354</td>
<td>.126</td>
<td>.144</td>
<td>.853</td>
<td>1.04</td>
<td>15</td>
<td>190.88</td>
<td>.419</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.251</td>
<td>.063</td>
<td>.067</td>
<td>.970</td>
<td>.73</td>
<td>8</td>
<td>140.00</td>
<td>.662</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.129</td>
<td>.067</td>
<td>.017</td>
<td>1.000</td>
<td>.40</td>
<td>3</td>
<td>71.00</td>
<td>.753</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Significance level *p<.05

Further interpretation was conducted to determine the relative importance of each of the variables in the canonical correlation. Table 2 shows the relative importance of each individual variable to the first canonical correlation. The first canonical variable for the emotional intelligence variables had a weighted difference of EI-A (-.964), EI-E (.453), EI-F (-.258), EI-G (.401), EI-C (-.212), and EI-B (.164), with more emphasis on EI-A. The canonical cross-loading of the variable EI-A (.410) also had the highest correlation with the canonical variable of the amount of interaction variable and the same sign as the canonical weight. Some of the remaining variables appear to have high weights. However, their contribution was not worthy to consider, since their canonical cross-loadings were lower than .30 and their squared multiple correlations indicate that they have no predictable power for the canonical correlation.

The first canonical variable for the amount of interaction variables had a weighted difference of MT (1.333), WA (-.957), MA (.581), and WT (-.119), with more emphasis on the variables of MT, WA, and MA. However, while the canonical cross-loading of the variable MT was .318, the canonical cross-loading of the variable WA was a mere .093 with a sign opposite from its canonical weight. This result might indicate that the variable WA is not actually important in the canonical correlation. Rather, MA looks to be the important variable in the canonical correlation in
that its value of cross-loading is .339. This result is consistent in the squaring of multiple correlations. The squared multiple correlations show that the first canonical variable of the amount of interaction has some predictive power for the number of messages in the synchronous text interaction (.101) and the number of messages in the synchronous audio interaction (.115).

Table 2. Relative Importance of Variables for the First Canonical Correlation between EI and the Amount of Interaction

<table>
<thead>
<tr>
<th>Variate/Variables</th>
<th>Canonical Weights</th>
<th>Canonical Cross-Loading</th>
<th>Squared Multiple Correlations</th>
<th>Redundancy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>1.333</td>
<td>.318</td>
<td>.101</td>
<td>.068</td>
</tr>
<tr>
<td>WT</td>
<td>-.957</td>
<td>.093</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>.581</td>
<td>.339</td>
<td>.115</td>
<td></td>
</tr>
<tr>
<td>WA</td>
<td>-.119</td>
<td>.215</td>
<td>.046</td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td>.044</td>
</tr>
<tr>
<td>EI-A</td>
<td>-.964</td>
<td>-.410</td>
<td>.168</td>
<td></td>
</tr>
<tr>
<td>EI-E</td>
<td>.453</td>
<td>.004</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>EI-B</td>
<td>.164</td>
<td>-.080</td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td>EI-F</td>
<td>-.258</td>
<td>-.272</td>
<td>.074</td>
<td></td>
</tr>
<tr>
<td>EI-C</td>
<td>-.212</td>
<td>-.092</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>EI-G</td>
<td>.401</td>
<td>.082</td>
<td>.007</td>
<td></td>
</tr>
</tbody>
</table>

Relationship between social bond and interaction

The results showed a positive relationship between students’ social bond and the types of interaction in synchronous text interaction. This relationship was not found in asynchronous interactions and synchronous audio interaction. The results showed no significant canonical correlation between students’ social bond and the amount of interactions both in synchronous and asynchronous online learning environment. Table 3 displays the canonical partial correlations between the social bond variables and the types of interaction variables. The first canonical correlation between social bond and types of messages in synchronous text interaction was found to be statistically significant (r=.407, p=.037). The squared canonical correlation for the first dimension was .165.

Table 3. Canonical Partial Correlations between Social Bond and the Types of Interaction

<table>
<thead>
<tr>
<th>#</th>
<th>Canonical Correlation</th>
<th>Squared Canonical (R²)</th>
<th>Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)</th>
<th>Test of H₀</th>
<th></th>
<th></th>
<th></th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Eigenvalue</td>
<td>Cumulative</td>
<td>F</td>
<td>DF 1</td>
<td>DF 2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.407</td>
<td>.165</td>
<td>.198</td>
<td>.755</td>
<td>2.05</td>
<td>9</td>
<td>175.38</td>
<td>.037*</td>
</tr>
<tr>
<td>2</td>
<td>.241</td>
<td>.058</td>
<td>.061</td>
<td>.989</td>
<td>1.16</td>
<td>4</td>
<td>145.00</td>
<td>.331</td>
</tr>
<tr>
<td>3</td>
<td>.055</td>
<td>.003</td>
<td>.003</td>
<td>1.000</td>
<td>.22</td>
<td>1</td>
<td>70.00</td>
<td>.639</td>
</tr>
</tbody>
</table>

Note. Significance level *p<.05

Further interpretation was conducted to determine the relative importance of each of the variables in the canonical correlation. Table 4 shows the relative importance of each individual variable to the first canonical correlation.

Table 4. Relative Importance of Variables for the First Canonical Correlation between Social Bond and Types of Interaction

<table>
<thead>
<tr>
<th>Variate/Variables</th>
<th>Canonical Weights</th>
<th>Canonical Cross-Loading</th>
<th>Squared Multiple Correlations</th>
<th>Redundancy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>-1.034</td>
<td>-.005</td>
<td>.000</td>
<td>.076</td>
</tr>
<tr>
<td>Work</td>
<td>.808</td>
<td>.114</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>.966</td>
<td>.319</td>
<td>.102</td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td>.038</td>
</tr>
<tr>
<td>Peer</td>
<td>.296</td>
<td>.243</td>
<td>.059</td>
<td></td>
</tr>
<tr>
<td>Instructor</td>
<td>-.246</td>
<td>.151</td>
<td>.023</td>
<td></td>
</tr>
<tr>
<td>Program</td>
<td>.970</td>
<td>.383</td>
<td>.147</td>
<td></td>
</tr>
</tbody>
</table>
The first canonical variable for types of interaction variables had a weighted difference of social type (-1.034), work type (.808), and management type (.966). All the variables seem to have an important contribution to the correlation. However, the canonical cross-loadings of the social type variable (r=-.005) and the management type variable (r=.114) were very small. Rather, the management type variable had the highest cross-loading (r=.319). This indicates that the management type variable was the most important variable in the canonical correlation, while the contribution of social type and work type variables to the relationship was small. Moreover, the squared multiple correlations show that the first canonical variable of types of interaction has some predictive power for management type interaction in the synchronous text interaction (.102). There was no predictable power for social type (.000) and work type (.013) interaction in the synchronous text interaction.

The first canonical variable for the social bond variables had a weighted difference of bond to peer (.296), bond to instructor (-.246), and bond to online program (.970), with greater emphasis on bond to program. The canonical cross-loading of the bond to program (.383) also had the highest correlation with the canonical variable of the social bond variable and the same sign as the canonical weight. Some of the remaining variables appear to have high weights. However, their contribution was not worthy to consider, since their canonical cross-loadings were lower than .30 and their squared multiple correlations indicate that they have less predictable power for the canonical correlation.

**Relationship between emotional intelligence and social bond**

No statistically significant canonical correlations between the emotional intelligence variables and the social bond variables were found. Four branch score variables, including EI-B1 (Perceiving Emotion), EI-B2 (Using Emotion), EI-B3 (Understanding Emotion), and EI-B4 (Managing Emotion) were input as a set of the emotional intelligence variables. For the social bond variables, peer, instructor, and program variables were input. Gender variable was also input as a control variable.

Seeing that there was no significant canonical correlation between the emotional intelligence variables and social bond variables, the Pearson partial correlations were further analyzed to determine the correlation between the task scores of emotional intelligence and social bond variables. According to the Pearson partial correlation, bond to peer was the only variable to be associated with the four emotional intelligence variables. Another canonical correlation analysis was conducted to further examine the relationship between the emotional intelligence variables and the bond to peer variable. Multiple regression analysis could be used to examine this one-dimensional relationship but canonical correlation analysis was used since canonical correlation also has an ability to do univariate multiple regression analysis and produce the same result as multiple regression analysis.

Table 5 shows the canonical partial correlation between the four emotional intelligence variables and the bond to peer variable. One canonical correlation was derived because the smaller set had one variable. The canonical correlation was found to be statistically significant (r=.349, p=.039) at the .05 alpha level. The squared canonical correlation for the first dimension was .129. This squared canonical correlation equals to R² in univariate multiple regression analysis.

<table>
<thead>
<tr>
<th>#</th>
<th>Canonical Correlation</th>
<th>Squared Canonical (R²)</th>
<th>Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)</th>
<th>Test of H₀</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Eigenvalue</td>
<td>Cumulative</td>
<td>F</td>
</tr>
<tr>
<td>1</td>
<td>.349</td>
<td>.129</td>
<td>.147</td>
<td>1.00</td>
<td>2.66</td>
</tr>
</tbody>
</table>

*Note. Significance level *p<.05

Further interpretation involves determining the relative importance of each of the variables in deriving the canonical relationships. Table 6 shows the relative importance of each individual variable to the significant canonical correlation.
The canonical variable for the emotional intelligence variable had a weighted difference of EI-B1 (.964), EI-B2 (-.317), EI-B3 (.332), and EI-B4 (.600). EI-B1 and EI-B4 appear to be important variables in the correlation. However, while the canonical cross-loading of the variable EI-B1 was .310, the cross-loading of the EI-B4 was not greater than .30. Moreover, the squared multiple correlations also indicate that EI-B1 has greater predictable power of the canonical correlation.

<table>
<thead>
<tr>
<th>Variate/Variables</th>
<th>Canonical Weights</th>
<th>Canonical Cross-Loading</th>
<th>Squared Multiple Correlations</th>
<th>Redundancy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Peer</td>
<td>1.000</td>
<td>.359</td>
<td>.129</td>
<td>.129</td>
</tr>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td>.040</td>
</tr>
<tr>
<td>EI-B1</td>
<td>.964</td>
<td>.310</td>
<td>.096</td>
<td></td>
</tr>
<tr>
<td>EI-B2</td>
<td>-.317</td>
<td>.145</td>
<td>.021</td>
<td></td>
</tr>
<tr>
<td>EI-B3</td>
<td>.332</td>
<td>.039</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>EI-B4</td>
<td>.600</td>
<td>.200</td>
<td>.040</td>
<td></td>
</tr>
</tbody>
</table>

In sum, there was a negative association between students’ emotional intelligence and the amount of interaction in the synchronous online environment. Students’ ability to perceive emotion by facial expressions and the number of text and audio messages were found to be critical factors in the relationship. Also, there was a positive relationship between students’ social bond and their types of interactions in the synchronous online environment. Students’ bonding to their online program and their management type interaction were found to be contributing variables in the association. Lastly, emotional intelligence with more emphasis on the ability to perceive emotion had a positive relationship with the bond to peer variable.

Conclusions and discussion

**Conclusion 1:** Students who have higher emotional intelligence have a greater degree of social bond in online learning. Specifically, individuals with a higher ability to perceive emotion tend to be more attached to peers in online learning environments.

The results indicate that students who have higher ability to perceive emotions have a greater sense of bond to their online learning peers. Previous literature supports the conclusion about the relationship between emotional intelligence and social bond in online learning. In traditional in-person relationships, students with a high ability to perceive emotions of self and others’ have better relationships with friends (Brackett et al., 2004). This positive relationship was found in the online learning environment in this study.

It was interesting that, while students’ ability to perceive emotion was found to be associated with their emotional attachment to peers, the other emotional abilities (i.e., understanding emotion, using emotion, and managing emotion) were found to have an insignificant contribution to the relationship. In traditional school settings, the ability to manage emotion was found to be associated with students’ social relationship with friends and school membership (Lopes et al., 2003; Zins et al., 2004). However, this relationship was not found in this study.

These inconsistent results may indicate the limited function of emotional intelligence in online learning environments. As Mayer et al. (2004) discussed, the four emotional abilities (i.e., perceiving emotion, using emotion, understanding emotion, and managing emotion) represent a hierarchy. In other words, they argued that individuals should be able to perceive their and others’ emotions in order to use, understand, and manage their emotion. The insignificant contribution of the other three emotional abilities in the relationship may result from this hierarchy of emotional abilities in online learning environments. Expressing one’s emotion or the ability to perceive others’ emotions in online environments is challenging (Wang & Reeves, 2007). Since emotional information processing in the emotion system (i.e., perceiving emotion) is challenged in online environments due to the limited emotional cues, the other emotional abilities (i.e., using emotion, understanding emotion, and managing emotion) may be challenged to help develop social bonds. The combination of the limited environmental characteristic of online learning and hierarchical representation of emotional abilities may reduce the association of using emotion, understanding emotion, and managing emotion with social bonding.
Another discussion point is the different nature of social bond in online learning environments. Even though social bond to peers was found to be an important variable, bond to instructor and bond to program were found to have little contribution to the relationship between emotional intelligence and social bond. This result may be due to the different qualities of bond. Wehlage et al. (1989) developed the social bond and school membership concept based on students’ attachment, commitment, involvement, and belief in schooling. While bonding to peers is based on their emotional attachment to peers, bonding to program seems to be based more on students’ belief and trust in the online program in terms of learning and professional development. This belief and trust of the program may be independent from their emotional ability, which may reduce the association of their bonding to their online program with their emotional intelligence.

The insignificant relationship of bond to instructor with emotional intelligence may be because bond to instructor is based more on rational commitment rather than emotional attachment. Wehlage et al. (1989) clarified that commitment is different from attachment in that commitment emphasizes the rational side of school membership, while attachment is the emotional aspect of school membership. Students commit to synchronous class participation, asynchronous discussions, and assignments in their online learning in order to achieve learning goals. Their commitment to the course activities affects instructors’ evaluation of their academic performance. So, they would commit to the course regardless of their emotional intelligence in order to earn good grades from the instructor. Students’ bond to instructor may be based on this commitment-oriented bonding, which may reduce the association of students’ bond to instructor with their emotional intelligence.

Conclusion 2: The inability to perceive emotions by facial expression brings a greater amount of interaction in synchronous online discussions. Students who cannot perceive emotions are willing to chat and talk, although not necessarily related to topics, during synchronous online classes.

This study provides an interesting finding that there is a negative relationship between students’ ability to perceive emotions by facial expression and their amount of synchronous online interaction. This may appear contradictory to studies in traditional classroom environments. Literature has shown that those who have higher emotional intelligence would have higher academic achievement in traditional classrooms (Brackett et al., 2004; O’Connor & Little, 2003; Trinidad & Johnson, 2001). Furthermore, Wang and Tucker (2001) found a strong positive correlation between students’ interactions in synchronous online environments and their academic achievement. These studies may suggest that individuals with a higher emotional intelligence may have higher levels of interaction and consequently higher academic achievement. However, in this study, emotional intelligence had a negative association with the amount of interaction in a synchronous online learning environment.

The limited environmental capacity to perceive emotions in online learning may bring greater emotional distance to students who have low ability to perceive emotions. Online learning environments have limited capacity to express one’s self and perceive others’ facial communication (Wang & Reeves, 2007), while traditional classroom environments provide students a greater opportunity to identify nonverbal cues such as facial expressions. It is not easy to perceive emotions in an online learning environment due to the emphasis on text-based communication, which does not require facial expression. It may be more challenging for individuals with a lower ability to perceive emotion to understand others’ feelings in online environments. For them, interactions by text typing and talking on the microphone may become alternative ways to overcome their inability to perceive emotion so that they are able to reach out to others. For individuals with a higher ability to perceive emotion, even several interactions may help them understand so that they do not feel they need further interactions.

Additionally, no relationship was found between emotional intelligence and types of interactions both in synchronous and asynchronous interactions. Students with higher emotional intelligence may have more social, work, and management type interaction, because emotional intelligence has been found to be positively associated with social relationship (Lopes et al., 2003; Mayer et al., 1999) and academic achievement in traditional classroom learning environment (O’Connor & Little, 2003; Trinidad & Johnson, 2002). However, this hypothesis was not supported in the current study.

Conclusion 3: Students who have stronger online program bonding have more management type interactions in synchronous discussions. Students’ belief regarding the online program legitimacy and efficacy may help them manage and facilitate discussions during synchronous sessions.
Literature in the field of online learning and school membership supports the conclusion of a positive relationship between social bond and interaction. Sense of belonging and community has been discussed as one of the important factors that affect students’ participation, engagement, and attitude in online learning (Harasim, 2000; McInerney & Roberts, 2004; Pools, 2000). The literature on school membership and effectiveness distinguished such emotional sense of belonging into bond to peer, teachers, and school and found their positive relationships with academic achievement and school engagement in traditional school environments (Newmann et al., 1992; Pryor, 1994; Wehlage et al., 1989). Considering the studies in both fields, it can be concluded that students’ social bonds are also associated with their online learning interactions.

In this study, students who had a higher degree of bonding to their online program had more management type of interactions during the synchronous sessions. This may indicate that adult online learners’ motivation to interact may result more from their professional interests and success in the online program than from their emotional attachment to peers and instructors. Adult online learners may have less need for peer acceptance and peer socializing. Most of them were employed and had family obligations. The purpose of their online learning was more focused on earning a degree, not necessarily on making friends through the online program. Thus, their interaction patterns during synchronous discussions may result from their belief and bonding to the online program, not merely attachment to peers and instructors.

Lastly, while management type interaction was found to be important, work type interaction was also found to be insignificant in the relationship. Since students are attached to the online program, they may want to perform well in the program so they tend to manage their learning by scheduling, facilitating, reinforcing group members, seeking for help, providing help, and conforming to the ways of online learning. Additionally, students’ work type interaction may be associated more with their prior knowledge in the subject matter and experience in the professions. Actually, all the types of interactions had significant and positive correlation with the age variable. After controlling the age variable, no relationship was found between work type interaction and student social bond. Even though the age variable does not represent students’ experience and knowledge, it may also mean that the older students, who may have more experience in the field of the subject, would tend to share their experience and knowledge and to work actively in the synchronous discussion regardless of their bonding.

**Recommendations**

Several recommendations are discussed in this section. First, the area of emotion has remained under-explored in understanding students’ interaction in online learning. This study brings awareness to the relationships between emotional intelligence, social bond, and interactions as emotional and social learning to the field of online learning. Interactions have been rarely discussed from an emotional dimension in the field of online learning. In fact, there are no empirical studies that investigate the role of emotional intelligence and social bond in students’ interactions in online learning environments. This study tried to provide a different lens in understanding online interaction by adopting an emotional-oriented perspective. Interaction is a defining factor in transactional distance between students and instructors in online learning environments (Moore, 1997). The more students have positive interactions with peers and instructors, the less they would have transactional distance between them (Moore, 1983, 1997). The inability to perceived emotions by facial expression may result in greater transactional distance (Moore, 1983), so students with a low ability to perceive emotions may have more interactions as found in this study. Transactional distance can be extended to students’ emotional distance functioned by interactions. Studies to examine online interaction from an emotional dimension can complement cognitive-oriented understanding of interactions in the field of online learning.

Second, another implication is to be aware of differences of emotional intelligence styles in different learning environments when conducting future studies. Online learning environments are different from traditional face-to-face learning settings. Even within online learning, synchronous online learning environments are different from asynchronous online learning settings. Results found in traditional face-to-face learning environments cannot be applied to online learning environments without evidence of empirical studies in online learning environments. For example, Fabio and Palazzeschi (2008) found that emotional intelligence and self-efficacy have a positive relationship in traditional school settings. Even though their study had a different design by using Bar-On’s (2004) model and was focused on high school teacher’s self-efficacy, the result showed a positive association between emotional intelligence and self-efficacy. Knowlton (2005) argued that embarrassment and low self-efficacy might
cause inactivity in asynchronous online discussions. Therefore, it might be possible to interpret that students with higher emotional intelligence may have higher self-efficacy and consequently have active interaction in asynchronous online discussions. However, this positive relationship between emotional intelligence and online interactions was not found, rather negative association was found in synchronous online interactions the current study. Hence, understanding the role of emotional intelligence in understanding students’ interactions in online learning environments should start with an awareness of different styles of emotional intelligence in different learning environments. Moreover, it is suggested to use other emotional intelligence models in future studies, since they may give different results in understanding the relationship between emotional intelligence, social bond, and interactions in online learning environments.

Third, future studies can examine what environmental cues can be used in order for students to express self emotions and perceive others’ emotions in synchronous learning environments. As Beaudoin (2002) discussed when students are quiet or have inactive typing, it may not mean that they are not listening or learning during synchronous sessions. Moreover, given the limited number of environmental cues in online learning, it is not easy to see whether students understand, are confused, or are bored during synchronous sessions. This limited ability to perceive others’ emotions may affect interactions. If there are any effective environmental cues by which students could easily perceive others’ emotions, their interactions and social relationship in online learning might be enhanced, and eventually the student online learning experience might be improved.

Finally, future studies should explore a more rigorous research design for asynchronous online environments to control courses, instructors, and classes that may affect the relationship to investigate the relationship between emotional intelligence, social bond, and interactions. In this study, no relationship was found in asynchronous interaction. This result might be because of different structures between synchronous sessions and asynchronous discussion. In synchronous sessions the instructor mainly talks. Students did not have any obligation or requirement to talk or chat, even though they can type anytime as they want. However, postings in asynchronous discussion were required and structured by assignments. Course structure has been found to influence students’ interaction in asynchronous interaction (Vrasidas & McIsaac, 1999). This may indicate that students’ amount of interaction is associated with course structure in asynchronous settings regardless of emotional intelligence or social bond. The influence of course structures, instructors, and cohorts may reduce the power to detect the association in asynchronous online interactions. More rigorous research design may produce different results for asynchronous interactions.

References


Wen-Chung Shih, Shian-Shyong Tseng*, Che-Ching Yang¹, Chih-Yu Lin and Tyne Liang¹
Department of Applied Informatics and Multimedia, Asia University, Taichung, 41354, Taiwan (R.O.C.) //
¹Department of Computer Science, National Chiao Tung University, Hsinchu, 30010, Taiwan (R.O.C.) //
sstseng@asia.edu.tw

*Corresponding author

ABSTRACT
In this work ubiquitous learning technologies are applied to Chinese scenic poetry appreciation. A folksonomy-based approach is proposed to accumulate knowledge about poems and their corresponding scenic spots. A user can use a “Tagging” operation by a smart phone to associate a concept (a word, a phrase or a sentence) with a scenic spot. These associations will be recorded in a folksonomy, which is then transformed into a Learning Guidance Tree to support personalized guidance in u-learning environments. The purpose of this work is to propose a folksonomy-based guidance mechanism. Also, we attempt to find students’ and teachers’ attitudes towards the innovative application through a case study. To show the effectiveness of the innovative approach, a ubiquitous learning system was developed to conduct Chinese scenic poetry appreciation activities for a fifth-grade Chinese course in Taiwan, in which forty-eight students participated. The survey results show that the system had a positive impact on students’ learning, especially on the affective domain including participation, motivation and interaction.

Keywords
Chinese scenic poetry, Ubiquitous learning, Folksonomy, Context-aware, Tagging

Introduction
In recent years context-aware ubiquitous learning (u-learning) has become a hot research topic (Chiou et al., 2010; Chu et al., 2010; Hwang, Kuo et al., 2010; Hwang, Tsai et al., 2008; Hwang et al., 2009; Yang, 2006). In context-aware ubiquitous learning environments, students’ learning behaviors can be detected by sensors embedded in real-world resources. Thus, learning systems can provide students with adaptive learning support. In spite of its benefits, context-aware ubiquitous learning has been primarily applied to the instruction of subjects with well-structured knowledge and subjects whose knowledge is mostly defined with correct answers, such as natural science, science experiments and language training. For example, digital guidance can be provided for pupils to observe and classify real-world plants and insects in ubiquitous learning activities (Hwang, Chu et al., 2010). The main reason for this is that existing approaches to acquiring knowledge of such subjects can be adopted for learning guidance. Repertory grids, decision trees and knowledge engineering approaches have been proposed to construct knowledge bases for learning guidance (Chu et al., 2010; Hwang, Chu et al., 2010). This study extends the application of ubiquitous learning to subjects which are aimed to train students to be creative and imaginative, beyond the cognitive skills of memorization and comprehension. Chinese scenic poems describe scenery in high-level ancient Chinese and express profound concepts by implication, which are difficult for children without related experiences to understand. Our idea is to use a ubiquitous learning environment to guide students to construct experience related to the poems.

In this paper a folksonomy-based approach is proposed to record, accumulate, organize and share students’ feelings related to poems while learning Chinese poetry outdoors. Then, folksonomy-based knowledge can be used to guide students in learning Chinese poetry. A folksonomy is a user-generated taxonomy employed to categorize and retrieve web content such as Web pages, photographs and Web links using user-defined labels called tags (Jaschke et al., 2008; Tsui et al., 2010). Typically, a folksonomy has a flat structure consisting of several user-defined categories. In essence, the tags are organized according to poems and locations. Also, what the students see, do and feel is recorded in the folksonomy. By making use of GPS-enabled (Global Positioning Systems) smart phones, the student’s location can be recorded by the backend server and thus, while learning Chinese poetry outdoors, students can receive appropriate recommendations and guidance which will enhance their learning experience and learning activities. In addition, students can express the intended concepts learned from the poems by providing text tags or/and photographs.
Chinese poetry instruction, unlike natural science instruction, is not structural and procedural. The value of its knowledge is not to request students to memorize standard answers. Instead, the instruction of Chinese poetry emphasizes thinking and feeling. Therefore, students need in-time guidance to transform experiences into knowledge. The aims of the research have been stated as the following research questions:

1. What are students’ attitudes toward the use of this system?
2. What are teachers’ attitudes toward the use of this system?
3. What is the effectiveness of the folksonomy-based guidance mechanism?

To show the effectiveness of the innovative approach a ubiquitous learning system was implemented to conduct Chinese scenic poetry appreciation activities in a fifth-grade Chinese course in Taiwan; forty-eight students participated in the learning activity. The results from the surveys and interviews that were conducted to understand the functionalities of the system and the learning effectiveness for the students show that the system had a positive impact on students’ learning, especially on the affective domain including participation, motivation and interaction. Based on the aforementioned idea a prototype has been implemented and surveys and interviews have been conducted. The results show that this proposed approach had a positive impact on students’ learning, especially on the affective domain including participation, motivation and interaction in Chinese poetry learning.

Preliminaries and Related Work

In this section, we review background knowledge about u-learning applications and Chinese poetry learning. Related research about this work is surveyed.

U-Learning Applications

The rise of u-learning results from the convergence of e-learning and ubiquitous computing (Lee, 2008; Ley, 2007). However, this topic is too new to have a widely accepted definition. Ogata et al. (Lyytinen & Yoo, 2002; Ogata & Yano, 2004) compare four learning environments according to two attributes: level of embeddedness and level of mobility. In their framework u-learning is characterized by high level of embeddedness and high level of mobility. A very broad-sense definition of ubiquitous learning is “anywhere and anytime learning”, no matter whether wireless communications or mobile devices are employed or not (Hwang, Tsai et al., 2008). According to many researches from this field (Hwang, Tsai et al., 2008; Ley, 2007; Lyytinen & Yoo, 2002; Ogata & Yano, 2004), ubiquitous learning contains mobile learning, pervasive learning and context-aware ubiquitous learning. In Si et al. (2006), u-learning applications were categorized as shown in Table 1, and a frame-based model was proposed to represent context-aware applications.

<table>
<thead>
<tr>
<th>Application Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location-aware learning guidance</td>
<td>Museum guide (Oppermann &amp; Specht, 1999)</td>
</tr>
<tr>
<td></td>
<td>Tour guide (Abowd &amp; Atkeson, 1997)</td>
</tr>
<tr>
<td></td>
<td>Conference assistant (Dey &amp; Futakawa, 1999)</td>
</tr>
<tr>
<td>Correlation-aware collaborative learning</td>
<td>Japanese teaching (Ogata &amp; Yano, 2004; Yin et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>Knowledge awareness map (El-Bishouty et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>P2P content access (Li et al., 2006)</td>
</tr>
<tr>
<td>Task-aware supervised learning</td>
<td>Requirement satisfied learning (Cheng &amp; Han, 2006)</td>
</tr>
</tbody>
</table>

Previous researchers have shown the benefits of mobile learning, pervasive learning and context-aware ubiquitous learning (Chiou et al., 2010; Hwang, Chu et al., 2010; Martín & Carro, 2009; Ogata & Yano, 2004; Pfeiffer et al., 2009; Wong et al., 2010; Yang, 2006). Subject learning has become popular in ubiquitous learning including natural science (Hwang, Chu et al., 2010; Pfeiffer et al., 2009; Yang, 2006), computer science (Martín & Carro, 2009) and language learning (Wong et al., 2010). Based on Bloom’s taxonomy (Bloom, 1956), the cognitive domain can be divided into six major levels: knowledge, comprehension, application, analysis, synthesis, and evaluation. Previous research is mainly focused on the bottom level, knowledge. However, this work aims to train students to be creative and imaginative and therefore extends the application of ubiquitous subject learning to the comprehension level.
To provide personalized (e-)learning, researchers have proposed several personalization mechanisms, such as adaptive presentation, adaptive navigation support, adaptive instruction and adaptive assessment (Chiou et al., 2010; Hwang, Chu et al., 2010; Hwang, Tsai et al., 2008; Hwang, Tseng et al., 2008; Martin & Carro, 2009). Among these personalization mechanisms, adaptive support has become popular in developing context-aware ubiquitous learning environment (Hwang, Tsai et al., 2008; Martin & Carro, 2009).

The project “Mudlarking in Deptford” in the UK (Sprake, 2009) bears a number of similarities to what we have done. It is a kind of guided tour that enables students to interact with the environment and other people using mobile devices. However, the purpose of this paper is to provide a tool that enables learners to experience a poem using surrounding contextual information. The difference is that students use surrounding context information to enrich Chinese poetry learning experience in a situated learning environment.

Chinese Poetry Learning

Chinese language learning has attracted world-wide attention in the recent decade (Bai & Li, 1999; Cai, 2003; Hung, 2007; Zhu, 2007). However, the flexible expression of Chinese makes it difficult to learn advanced features of Chinese, such as the poetry of the Tang Dynasty which is characterized by a lot of information embedded in a limited number of Chinese characters (Cai, 2008). Based on the theories of Situated Learning and Experiential Learning (Vince, 1998), outdoor instruction emerges as a promising way to enhance children’s Chinese poetry learning. When the learner is situated in an environment similar to that which the poem describes, it will be easier for the learner to understand what the poet meant, thus enhancing the learning performance. Applying u-learning to Chinese poetry learning is more challenging than applying u-learning to subjects with well-structured knowledge. As indicated by previous research, personalized guidance in u-learning activities is based on pre-defined knowledge structures, which are constructed by experts or machine-learning approaches in advance. For Chinese poetry learning, the primary goal is to inspire students’ feelings, affection and creativity, instead of memorization. Therefore, the knowledge to be conveyed to students cannot be defined in advance.

Folksonomy

Previous researchers have shown the benefits of applying annotation mechanisms to collaborative learning (Robert, 2009; Su et al., 2010; Yang, 2006; Yang et al., 2004; Yeh & Lo, 2009). In general, the annotation mechanisms can be classified into two types: taxonomic or non-taxonomic ones. Among these, non-taxonomic annotation mechanisms (i.e. arbitrary annotations) have become popular in Web 2.0-based annotation and tagging in collaborative learning environments. Through collaborative learning, folksonomic annotations can be built up into a taxonomy of annotations.

Folksonomy-based Guidance Mechanism

In this work, a folksonomy-based guidance mechanism is proposed to develop an outdoor learning environment for Chinese poetry instruction. With this mechanism, the learning system can guide the students to create experiences related to the poem by sensing their locations.

Learning guidance design and system framework

In this work we develop a learning flow for ubiquitous Chinese poetry instruction. To guide the students to appreciate poems during the situated learning process, we propose a folksonomy-based guidance mechanism. The learning flow consists of 4 stages.

Stage 1: Poem Recommendation
The system recommends to a student a list of poems, which are selected according to the student’s location, learning portfolio, etc. The location context is acquired by GPS. In this study the location of outdoor learning is by a lake, and there are two poems related to this location can be chosen by students: “Watching Fish” and “On Pond”, as shown in Figure 5. In this study we focus on reporting the results of “Watching Fish”. After the poem is selected by the
student, the teaching material for this poem is displayed on a smart phone. The student has to pass a quiz about this poem to enter the next stage.

**Stage 2: Sensory Experience**

During this stage, the smart phone prompts the student to note experiences related to this poem in this location using various senses, such as sight and hearing. The students are required to record their experiences by text or photos using their smart phones. If the student can not find anything to experience, the system will prompt other students’ tagging in the folksonomy to her/him. The student has to annotate at least one tag about this poem to enter the next stage.

**Stage 3: Action Experience**

During this stage, the smart phone prompts the student to do something related to this poem in this spot. The students are required to record their action experiences by text or photos using their smart phones.

**Stage 4: Reflection Tagging**

After sensory and action experience, the system prompts the student to tag her/his reflections about this poem. The goal of this stage is to help students organize their knowledge formed in the field-trip learning process.

During the learning stages, students are guided according to a learning guidance tree, which is derived from a folksonomy. The overview of the folksonomy transformation process is illustrated in Figure 1. In Phase 1 students tag their opinions related to the poems and categorize their tags into the existing folksonomy. For example, a student can assign a tag “fishing” to the category “recreation”, which means “fishing” is a kind of “recreation”. A folksonomy consists of a number of categories. Students can assign their tags, generated during the learning process, to one of these categories on their own judgment. In Phase 2 experts and teachers collaboratively review the folksonomy and transform it into a learning guidance tree in a wiki-based manner. In this work the transformation process is executed by humans including students, teachers and experts.

![Figure 1. The process of folksonomy transformation](image)

**Example**

The poem “Watching Fish”, as shown in Figure 2, is chosen by the student. The corresponding learning guidance tree is then constructed, as shown in Figure 3.

![Figure 2. The poem recommended by the system (with translation in English)](image)
The proposed system is designed to guide students to learn in a step-by-step manner. The benefits from this system will be reduced if the learners do not follow the learning flow. The learning field and screenshots are shown in Figure 4 and Figure 5, respectively.

![Figure 4. Learning around the lake](image)

![Figure 5. Screen shots of learning guidance during Stage 2](image)

**Research results and Discussion**

In this work, the ubiquitous learning environment of the Chinese poetry course is a lake located in a university. There are lotuses and fish in the lake. The "Chinese poetry" course is the school-based curriculum of the chosen elementary school. The complete teaching cycle lasted for eight weeks, two hours per week. Students are required to learn two Chinese poems during this period. The proposed system is implemented with the aim to conduct Chinese scenic
poetry appreciation activities in context-aware ubiquitous learning environment so that both students’ and teachers’ attitudes toward the use of system were investigated.

The folksonomy used in this work is collected by the Chinese Poetry Learning Website (http://140.113.167.118:8080/intention2). We obtained a subset of tag data by selecting all tags from users who have participated in this experiment during an eight week period. Statistics of the datasets are given in Table 2.

Table 2. Dataset Statistics

<table>
<thead>
<tr>
<th>Poem Class</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of users</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>No. of tags</td>
<td>29</td>
<td>37</td>
</tr>
<tr>
<td>No. of Categories</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

These two questionnaires are originally developed by Hwang et al. (2010) and are modified to evaluate the attitudes of teachers and students towards the proposed system. The questionnaire for teachers includes twelve questions and the questionnaire for students includes twenty-six questions. The data collected from the questionnaire are measured using a five-point Likert scale that ranges from, *strongly disagree* (1-point) to *strongly agree* (5-point). These questions can be classified into four dimensions: perceived usefulness, perceived ease of use, learning satisfaction and willingness for future use (Davis, 1989). The researchers utilized Cronbach’s alpha analysis to evaluate the internal consistency of each dimension of this questionnaire. The Cronbach alpha values in four dimensions are higher than 0.70. This implies the reliability of the questionnaires is sufficiently high (Wortzel, 1979). During this questionnaire design process, the questions were verified and validated by domain experts (five professors). Some ambiguous or unsuitable questions were modified, removed, altered, or arranged in a proper order. The design of the questionnaires highly relates to the two research questions. The questionnaire for teachers corresponds to the former question; the questionnaire for students corresponds to the latter question.

The questionnaire for students was pilot tested with 24 students different from the 48 subjects. Completed scales were received from all participants. These students are asked to mark any problems on the questionnaire, such as misleading worded questions, improper order of questions, improper negations of questions, or if it takes excessive time to finish this survey. Data were then analyzed using SPSS to assess the reliability and validity of the questionnaire. The questionnaire is revised according to participants’ feedback before using it in the study.

For the course of “Watching Fish” twenty-nine distinct tags were generated. These tags were then analyzed by experts and categorized into four categories: cherishing animals (10 tags), education (3 tags), recreation (14 tags) and others (2 tags). After checking the system log, we found that more than half of the students (15/24) tended to reuse and refine extant tags to generate their own tags. These tags mainly fall into two categories: cherishing animals (10 tags) and recreation (14 tags). This might mean the two categories represent the consensus of students about this poem. This is to say, they have the common impressions of cherishing animals and recreation for this poem.

We have conducted semi-structured interviews to provide deeper insight into participants' views in the study. We have interviewed all of the students in the experiment group and teachers to acquire detailed feedback.

- The feedback from high achievement students shows their interests in the proposed learning activities. One of the students said, “It is interesting to use a smart phone to learn in an authentic environment.” Another student said, “Using photos to annotate poems is cool.” Still another student said, “It is exciting to create a new tag which has not been posted by other students.”
- The feedback from low achievement students shows their learning motivation has been elevated. One of the students said, “It is helpful for me to browse annotations made by others.” Another student said, “With the aid of learning guidance, I realized how to use tags to express personal feelings.”
- The feedback from teachers shows the proposed approach can assist students to appreciate the poem. One of the teachers said, “Everyone may have dissimilar feelings toward the same poem. Through the proposed approach, students can share their tags.” Another teacher said, “Through the activity of annotation sharing, students can learn other students’ experiences regarding the poem.”

In the first evaluation, a survey concerning the u-learning system performance, nine elementary school teachers who had experience teaching Chinese poetry courses were asked to fill in a questionnaire about the system functionalities and interface designs (as shown in Table 3). Note that this survey focused mainly on the use of smart phones, not on the folksonomy creation algorithm. These nine teachers have between six and nine years experience of working in
elementary schools. Thus they are both experienced in teaching and guiding students through the Chinese poetry courses.

Table 3. Statistical results of the questionnaire from nine teachers

<table>
<thead>
<tr>
<th>No.</th>
<th>Items</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>This teaching model can elevate learning motivation.</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.56</td>
</tr>
<tr>
<td>2</td>
<td>This teaching model can elevate learning effectiveness.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.22</td>
</tr>
<tr>
<td>3</td>
<td>This teaching model can help students to express what they think about the poem.</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.44</td>
</tr>
<tr>
<td>4</td>
<td>The system has high feasibility.</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.11</td>
</tr>
<tr>
<td>5</td>
<td>The system can help teachers to reduce their teaching burden.</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4.00</td>
</tr>
<tr>
<td>6</td>
<td>I would recommend this system to my colleagues.</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4.11</td>
</tr>
<tr>
<td>7</td>
<td>The use of icons and text in this system is coherent.</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.33</td>
</tr>
<tr>
<td>8</td>
<td>I can easily find the functions in the system.</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3.67</td>
</tr>
<tr>
<td>9</td>
<td>Switching between pages in the system is easy.</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3.89</td>
</tr>
<tr>
<td>10</td>
<td>The interface design including text color, size, and icons is easy to identify.</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4.56</td>
</tr>
<tr>
<td>11</td>
<td>The messages provided by the system are brief and concise.</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.33</td>
</tr>
<tr>
<td>12</td>
<td>The overall manipulation of the system is easy.</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4.33</td>
</tr>
</tbody>
</table>

As shown in Table 3, most of the teachers believe that this system is suitable for Chinese poetry learning and think that the system can promote students’ motivation and learning performance (Items 1-3). However, in terms of practical application to instruction, some teachers express their concern and hesitation (Items 4-6). Having experienced this system through smart phones, the feasibility of the system is acceptable for most teachers. However, some teachers think the application of this system requires teachers to prepare in advance, thus increasing their work load.

In terms of the system interface, most teachers think this system is user friendly (Items 7 and 11). However, some teachers express neutral attitudes towards usability of the system (Items 8, 9, 10 and 12). According to interviews with these users, the main reason is that they are not used to operating mobile devices with small displays. Also, they think that some default functions of the smart phone are not so obvious to ordinary users.

The second evaluation addresses the students’ responses to this experiment. The results are shown in Table 4, and discussed as follows. The first five items address the learning behavior and attitude of students before the experimental class. Most students have experience of using a digital camera, and leaning Chinese poetry using photos and graphics as supplementary tools (Items 3 and 5). However, some students have no experience of using smart phones, learning Chinese poetry outdoors, and thus have difficulty in expressing what they feel about a poem (Items 1, 2 and 4). The results show that the proposed approach is novel to most students. Also, most students have enough information literacy to conduct the innovative course.

The next nine items (Items 6 – 14) address the learning motivation and progress of students during the experimental class. Most students like to learn Chinese poetry outdoors because they are interested in outdoor activities (Item 6), smart phones (Item 7), or Chinese poetry (Item 8). Furthermore, students think that the proposed approach is more interesting than traditional classes (Item 9).

Table 4. Statistical results from the questionnaire for forty-eight students

<table>
<thead>
<tr>
<th>No.</th>
<th>Before class</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I have experience in learning Chinese poetry outdoors.</td>
<td>9</td>
<td>16</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td>3.04</td>
</tr>
<tr>
<td>2</td>
<td>I have used a smart phone before.</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>13</td>
<td>12</td>
<td>2.65</td>
</tr>
<tr>
<td>3</td>
<td>I have used a digital camera to take photos outdoors before.</td>
<td>21</td>
<td>20</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4.29</td>
</tr>
<tr>
<td>4</td>
<td>I can easily express what I feel about a Chinese poem.</td>
<td>7</td>
<td>17</td>
<td>11</td>
<td>7</td>
<td>6</td>
<td>3.25</td>
</tr>
<tr>
<td>5</td>
<td>The teacher would use supplementary tools such as photos and graphics to teach Chinese poetry.</td>
<td>25</td>
<td>18</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>4.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>During class</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>I like to learn Chinese poetry outdoors because outdoor activities are interesting.</td>
<td>9</td>
<td>22</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>3.75</td>
</tr>
</tbody>
</table>
I like to learn Chinese poetry outdoors because I like smart phones.  
I like to learn Chinese poetry outdoors because I can learn more about it.  
The smart phone’s guidance when learning Chinese poetry outdoors is more interesting than the teacher’s classes.  
I tried to use the smart phone to search for related information during the outdoor activities.  
I like learning with the smart phone’s guidance.  
I think using a smart phone is a convenient way to learn.  
I think using a smart phone is easy to learn.  
I am more willing to learn and discuss with peers when using a smart phone.  

<table>
<thead>
<tr>
<th>No.</th>
<th>After class</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Being able to use a smart phone is one of the benefits of this class.</td>
<td>18</td>
<td>21</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>4.19</td>
</tr>
<tr>
<td>16</td>
<td>Being able to use a smart phone for learning is one of the benefits of this class.</td>
<td>20</td>
<td>24</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.33</td>
</tr>
<tr>
<td>17</td>
<td>I know more about the poems after taking the class.</td>
<td>17</td>
<td>24</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4.21</td>
</tr>
<tr>
<td>18</td>
<td>I am more concerned about the things around me related to Chinese poetry after taking the class.</td>
<td>21</td>
<td>20</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4.29</td>
</tr>
<tr>
<td>19</td>
<td>I would spend more time after class to learn about Chinese poetry.</td>
<td>14</td>
<td>20</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>3.90</td>
</tr>
<tr>
<td>20</td>
<td>I think using a smart phone to learn is more interesting than before.</td>
<td>26</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.50</td>
</tr>
<tr>
<td>21</td>
<td>I am happy to use a smart phone in the Chinese poetry learning process.</td>
<td>27</td>
<td>19</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.52</td>
</tr>
<tr>
<td>22</td>
<td>I think a smart phone, which can provide one-to-one teaching, can help me to learn better than the teacher’s explanations.</td>
<td>22</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4.42</td>
</tr>
<tr>
<td>23</td>
<td>I have gained a lot from learning by using a smart phone.</td>
<td>25</td>
<td>17</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>4.40</td>
</tr>
<tr>
<td>24</td>
<td>Learning with a smart phone makes me feel more comfortable.</td>
<td>19</td>
<td>22</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4.25</td>
</tr>
<tr>
<td>25</td>
<td>Learning with a smart phone can enhance my learning motivation.</td>
<td>22</td>
<td>22</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.38</td>
</tr>
<tr>
<td>26</td>
<td>I would recommend the smart phone learning system to others.</td>
<td>19</td>
<td>18</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>4.10</td>
</tr>
</tbody>
</table>

In terms of functions, some students do not use smart phones to get related information (Item 10). The possibilities include that the folksonomy-based scaffolding system has already embedded sufficient expert knowledge to guide their learning, and that the smart phone interface is too small to allow comfortable searching for external information. From item 11 to item 14, it can be seen that students enjoy the learning guidance and prompt, and that the system is convenient and easy to use. Therefore, they like to use this system to learn with their peers.

The final twelve items address students’ attitudes toward the proposed approach. The results show that students think that using smart phones is beneficial for learning Chinese poetry (Items 15 and 16). Most students think their learning performance is raised through the class (Items 17 and 18). Also, most students agree that the experience of learning Chinese poetry using smart phones outdoors is interesting, helpful and comfortable (Items 20 - 25). Hence, they would recommend the u-learning system to others (Item 26). However, some students are not willing to spend more time after class to learn about Chinese poetry (Item 19).

The students’ responses to the questionnaire items before and after participating in the u-learning activity led to some interesting results. As shown in Table 5, we found that most students did not have prior experience of using a smart phone (mean = 2.65 for Item 2), but after the u-learning activity, they stated, “Learning with a smart phone makes me feel more comfortable” (mean = 4.25 for Item 24); moreover, the students felt that “Learning with a smart phone can enhance [their] learning motivation” (mean = 4.38 for Item 25). The lack of previous experience of using a smart phone did not affect the learning motivation of the students.
Table 5. Descriptive statistics for selected pre- and post-class questionnaire items

<table>
<thead>
<tr>
<th>No.</th>
<th>Questionnaire item</th>
<th>Mean</th>
<th>n</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Before class</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I have used a smart phone before.</td>
<td>2.65</td>
<td>48</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td><strong>After class</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Learning with a smart phone makes me feel more comfortable.</td>
<td>4.25</td>
<td>48</td>
<td>0.70</td>
</tr>
<tr>
<td>25</td>
<td>Learning with a smart phone can enhance my learning motivation.</td>
<td>4.38</td>
<td>48</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Regarding item 15, the students who stand neutral toward using smart phone to learn have the following opinions. Five students are not used to learning with a smart phone. Two students have difficulties in operating the system. Regarding item 17, the students who stand neutral have the following opinions. Four students think they would acquire more knowledge by reading a textbook than that acquired by tagging. Regarding items 18 and 19, the students who stand neutral have the following opinions. Four students have no interest in poems. Therefore, they do not care about the things around them related to Chinese poems and they would not spend time learning Chinese poems after class. Three students think that learning poetry is interesting but they have other things to do. So they would not spend more time after class to learn about Chinese poetry.

**Discussion**

Based on the status of u-learning research mentioned in the Introduction, the findings can be interpreted as follows:

- It is important to consider a guidance mechanism when designing u-learning environments. The proposed folksonomy-based teaching model can provide students with appropriate guidance during their outdoor learning process, as also indicated in Hwang et al. (2010). In addition, the integration of ubiquitous learning into instruction can promote students’ learning motivation. This viewpoint can also be found in other subject matters, including mathematics, natural science, chemistry experiments (Chiou et al., 2010; Chu et al., 2010; Hwang, Chu et al., 2010; Hwang, Kuo et al., 2010).
- Regarding Chinese poetry learning, we found that the right guidance, obtained by utilizing other students’ tags, can help students think more about the relationships of poems to their surroundings. In this work we see the effect of ubiquitous learning Chinese poems outdoors. We found that applying u-learning technologies to Chinese poetry learning to achieve higher educational objectives is feasible. Existing research of u-learning focuses on a limit set of subjects with learning activities of observation and classification. More innovative applications of u-learning technologies to other subjects can be explored.

With the aid of folksonomy-based guidance mechanism, high achievement students show their great interests in the proposed learning activities. Low achievement students indicate this mechanism can help them understand how to annotate, help them think, and further help them understand poems. Teachers also have a positive attitude toward the folksonomy-based guidance mechanism.

One alternative guidance mechanism in u-learning environments is the decision-tree-oriented approach which has been developed for learners to observe and classify real-world plant in natural science courses (Hwang, Kuo et al., 2010). The strategy of constructing the decision tree is to employ a machine-learning algorithm for deriving decision rules from a set of training examples. However, this approach does not work when training examples are not available. This is the reason why we propose the folksonomy-based approach to construct a guidance structure for Chinese poetry learning.

Web-based discussion board systems are alternatives in Chinese poetry learning. Most of web-based discussion board systems support the creation and sharing of comments. Wikipedia is a well-known example in which the learners collaboratively edit the online content. This mechanism also allows users to add comments to an online teaching material and tags noted by other users would form a folksonomy. In spite of the aforementioned strengths, it is not easy to form consensus among users of a general-purpose web-based discussion board system (Hardaker & Smith, 2000; Marshall & Brush, 2002). In contrast, the proposed mechanism attempts to provide guidance during students’ learning process. Default tags provided by teachers can serve as a start point of think about the poem.

The pedagogical and instructional design theories used in this study include Kolb’s experiential learning theory (Kolb, 1984), situated learning and Scaffolding. Based on Kolb’s theory, we have proposed a four-stage learning
flow: poem recommendation, sensory experience, action experience and reflection tagging. The third stage “action experience” is designed according to Dewey’s “learning by doing” theory (Dewey, 1938) which emphasizes the effect of action while learning. Based on Wood et al.’s (1976) definition of scaffolding and Vygotsky’s (1978) model of instruction, the proposed four-stage learning guidance mechanism can be seen as assistance from more knowledgeable peers and teachers that enables children to achieve what is beyond their capability to complete by themselves. Therefore, teachers can help learners to appreciate Chinese poems within their zones of proximal development (Vygotsky, 1978). Situated learning emphasizes the value of acquiring knowledge in both social and situational contexts (Brown et al., 1989). Based on situated learning theory (Billett, 1996; Lave & Wenger, 1991; Young, 1993), we propose to learn Chinese poems outdoors with the help of the proposed u-learning guidance mechanism.

The weakness or limitation of the proposed system is the granularity of context information. Since this work adopts “coarse-grained” context information (GPS information), instead of “fine-grained” context information (RFID information), the proposed system cannot acquire real-time interaction from environments. However, using GPS information results in easy deployment and application of u-learning environments compared with previous u-learning studies.

Conclusions

Chinese scenic poems describe scenery in high-level ancient Chinese, which is difficult for children to understand by vocabulary explanation and recitation. Context-aware ubiquitous learning outdoors, which combines real-world and cyber-world resources, is a promising way for children to learn Chinese scenic poetry by experiencing real scenery. However, most existing ubiquitous learning systems are based on experts’ efforts to construct the knowledge and content for learning guidance. Our idea is to propose a folksonomy-based approach which can accumulate the knowledge about poems and their corresponding scenic spots. A user can use a “Tagging” operation by a smartphone to associate a concept (a word, a phrase or a sentence) with a scenic spot. These associations will be automatically integrated into a folksonomy to support a u-learning system. To show the effectiveness of the innovative approach, a context-aware ubiquitous learning environment was implemented for the “Chinese poetry” unit of a fifth-grade Chinese language course in Taiwan, and forty-eight students participated in the learning activity. From the surveys and interviews that were conducted to understand the functionalities of the system and the learning effectiveness for the students, the results show that the system had a positive impact on students’ learning, especially on the affective domain, including participation, motivation and interaction.

From the survey and interviews with the teachers who experienced the innovative learning environment, we saw that this environment was highly accepted by most of the teachers. They agreed that, with the folksonomy-based scaffolding in the outdoor learning environment, students can better learn about experiences related to the poems; therefore, it was suggested that more learning activities for other courses that aim to enhance students’ ability and desire to think about the subject matter can be conducted in this test field. In addition, we observed that the folksonomy-based guidance mechanism integrated in the course seems to be effective in enhancing students’ learning motivation, especially with new tools and innovative teaching methods. The statistical data show that, no matter whether students have used a smart phone before or not, they think using one to learn is not only interesting, but also helpful. The students also indicated that their learning motivation is high thanks to the use of smart phones, including their convenience and ease of use, and guidance with the folksonomy-based scaffolding. Therefore, it is worth conducting long-term and large-scale experiments in the future to observe the learning effectiveness of such an innovative approach.

Acknowledgements

This research was partially supported by National Science Council of Republic of China under the number of NSC97-2511-S-468-004-MY3, NSC98-2511-S-468-004-MY3, NSC98-2511-S-468-002, NSC99-2511-S-468-003 and NSC99-2221-E-009-130-MY2.
References


How Concept-mapping Perception Navigates Student Knowledge Transfer Performance

Kuo-Hung Tseng¹*, Chi-Cheng Chang², Shi-Jer Lou³, Yue Tan¹ and Chien-Jung Chiu¹

¹Meiho University, Taiwan // ²National Taiwan Normal University, Taiwan // ³National Pingtung University of Science & Technology, Taiwan // gohome8515@gmail.com // samchang@ntnu.edu.tw // lousj@ms22.hinet.net // yuetan@gmail.com // sea-guy@yahoo.com.tw

*Corresponding author

ABSTRACT

The purpose of this paper is to investigate students’ perception of concept maps as a learning tool where knowledge transfer is the goal. This article includes an evaluation of the learning performance of 42 undergraduate students enrolled in a nanotech course at a university in Taiwan. Canonical correlation and MANOVA analyses were employed to examine if students’ perceptions toward concept mapping have a positive relationship with knowledge transfer; that is, students who perceive concept mapping more positively tend to perform knowledge transfer better than those who perceive concept mapping less positively. The results revealed that positive concept-mapping perception is helpful for knowledge transfer in five learning stages: acquisition, communication, application, acceptance, and assimilation.

Keywords

Concept mapping, Knowledge transfer, Knowledge management, Nanotech

Introduction

In recent years, the development of information technology has led to rapid knowledge flow. People gain knowledge through a variety of ways. Consequently, knowledge management (KM) becomes increasingly more important for individuals to understand what information is essential, how to administer this essential information, and how to transform essential information into permanent knowledge.

Knowledge transfer (KT) is an important aspect derived from KM. KT takes place to the extent that an organization benefits from knowledge acquired at other organizations or in other parts of the same organization, and can improve the performance level of the organization (You, Li, & Yu, 2006). In practice, KT has been applied in different fields such as business (Darr & Kurtzberg, 2000) and biology (Schönborn & Bögeholz, 2009). KT, in general, is when experts in a field of study transmit their knowledge to younger generations. In a like manner, the transfer would also occur among people of the same generation. Our study focuses on KT in an educational context; in particular, we examine the impact of concept mapping (CM) on KT in a science lecture.

It is widely acknowledged that concept maps (CMs) have been used in many facets of education and training. CM has many advantages, including playing a multi-level tool role, scaffolding for cognitive processing, summarizing and organizing what has been learned, supporting collaboration, consolidating educational experiences, teaching critical thinking, improving achievements and interests of learning, etcetera (Adamczyk, Willison, & Williams, 1994; Chiou, 2008; O’Donnell, Dansereau, & Hall, 2002). Additionally, numerous studies indicate that CM is an effective learning strategy that precipitates meaningful learning for different learners in a variety of fields, such as science (Kinchin, 2001).

While much research has explored the relationship between CM and learning achievements, little research has focused on KT. Therefore, we investigated how students use CMs to achieve KT in an effort to provide valuable reference work for science education.

Literature review

The role of advanced organizers (AO) and graphic organizers (GO) in CM learning

Assimilation theory addresses the issue of using CMs to promote meaningful learning. The theory also explains how concepts might be accurately acquired and organized within a cognitive structure in learners through the use of a
variety of teaching/learning strategies (Novak, 2003). Ausubel (2002) advocates the use of advanced organizers (AO) to foster meaningful learning, and he describes the role of AO in the progressive differentiation of learned concepts. Ausubel further states that the major goal of schooling is to foster development of schemas that learners can use to acquire other relative information and to stimulate integrative reconciliation in the process of acquiring knowledge. Through AO, integrative reconciliation occurs in the consciousness of learners who perceive that learned concepts are related or not related.

In order to improve meaningful learning, the content to follow should be well organized. The schema production should be under guided advice of teachers. It requires teachers and students to monitor their thinking and learning in traditional classroom settings. However, the advent of various multimedia in which a broad range of possible elements can be used has given rise to modern AO in computer-based or web-based learning contexts (Brabec, Fisher, & Pitler, 2004; Langan-Fox, Platania-Phung, & Waycott, 2006; Mayer, 2003; Vekiri, 2002). Graphic organizers (GO) form a powerful visual picture of information and allows the mind to “see” patterns and relationships. GO can help motivate, increase recall, assist understanding, create interest, combat boredom, and organize thoughts (Robinson & Kiewra, 1995). AO, along with GO, appear to be constructed in the form of text passages, graphical representations, and maps, which are commonly used in some commercial computer applications, such as FreeMind. AO and GO have been used successfully in a wide range of courses and knowledge domains.

**Relationship between CM perception and KT performance**

Pinto (1997) explains that CM perceptual strategy or pedagogy improves the KT performance because students facilitate the application of knowledge. They use CM perceptual strategy to gain insight into new and existing knowledge. CM can help students develop good learning habits, which may contribute to future KT.

Some studies found that CM generally has positive effects on both students’ achievements and attitudes and promotes meaningful learning by helping them see the links between scientific concepts (Adamczyk, Willison, & Williams, 1994; Chiou, 2008). The benefits of CM reported in the literature include organizing information, assisting learning, communicating particularly complex CMs, understanding literature, improving clarity, testing students’ learning progresses, and more (Novak & Gowin, 1984; Novak, Gowin, & Johansen, 1983; Ruddell & Boyle, 1984). For these reasons, it is argued that CMs stimulates KT.

CM is also useful for knowledge retention and transfer on learning nuclear technology (Mann & LeClair, 2009). González, Palencia, Umaña, Galindo, and Villafrade’s (2008) study in medical physiology students showed that CM strategy promoted meaningful learning that allowed the students to transfer this knowledge to solve problems. In addition, CM strategy had a greater impact on the students who came into the study with the lowest cognitive competence. So, educators have recognized that the process of creating CMs is important for meaningful learning. It is proposed, based on the above literature, that perceived CM also promotes the conversion of the stages of KT performance.

**The process of KT**

Gilbert and Cordey-Hayes (1996) identified five stages in the process of KT, including acquisition, communication, application, acceptance, and assimilation. At the organizational level, KT has to go through these five active learning stages to achieve an organizational goal, as shown in Figure 1. In this study, we define KT as the interactive process between individual learning and organizational level learning to continually develop, change and react to internal and external effects to achieve competitive success.

The process of KT is dynamic and continuous. Individuals complete this process by going through the five stages. KT is both the movement of knowledge from one place to another and the altering of that knowledge into an understandable form (Major & Cordey-Hayes, 2000). For this reason, we use Gilbert’s model to illustrate how CM affects KT through the five stages.
Acquisition. Knowledge acquisition is the first step of KT. Before the knowledge is able to be transferred, it has to be acquired. A student might learn from various sources: direct experience, mass media, schools, parents, and so on. Sources of knowledge influence the type of knowledge acquired.

Novak and Gowin (1984) claimed that CMs can be viewed as a schematic device for representing a set of concepts and meanings embedded in a framework of propositions, as a tool for students to examine their prior knowledge. Therefore, CM is a particularly good way of organizing information related to a problem or subject. The construction of CMs help us pull together information already known about a subject while integrating new information as we learn and expand our understanding (Akinsanya & Williams, 2004). Some studies also propose that CM provide learners with a key scaffold to help them relate concrete examples to the conceptual structures and help retain the information to be learned through graphic formats (Reader & Hammond, 1994; Adamczyk, Willison, & Williams, 1994).

Accordingly, we considered that CMs help students understand new knowledge. Thus, the stage of acquisition in KT occurs. Students’ acquisition of new knowledge is often affected by their past knowledge and sources of new knowledge.

Communication. Once the knowledge is acquired, communication follows. Communication increases knowledge and understanding, which enriches prior knowledge and stimulates new thoughts. Individuals use this process to enhance the effectiveness of KT. Thus, fluent communication ought to be established in advance in order to lead the whole model to be efficient. In particular, communication mechanisms and barriers need to be identified to encourage KT. KT is a process through which the individuals are affected by another experience produced by knowledge. And knowledge can be transferred by moving tools and technologies, routines, and networks. Any transfer does not exist in one party. The movement should occur between a source, where the knowledge is acquired from and a recipient, to whom the knowledge is transmitted. Learners often share and transmit the acquired knowledge to others who do not possess this knowledge. This kind of learning approach is similar to a collaboration style. Many studies present collaboration as a result of successful interaction between parties. The students were able to engage in a discussion to consider the knowledge context (Kao, Lin, & Sun, 2008; Kwon & Cifuentes, 2007). Studies argue that CM can be used to display individual knowledge structures for comparison at different stages of the learning process, and to distinguish between expert and novice knowledge structures (Kinchin, 2001). Furthermore, Cicognani (2000) claimed that CM is viewed as a collaborative strategy, is applied to learning communication, and is described as a process through which one or more participants use brainstorming techniques. Consequently, CM is a reflective process in which students engaged in a rich discussion about how and why the concepts are related.

CM helps students not only understand the relationships among concepts, but also communicate this understanding to their classmates. Through discussion, individuals could understand better what they had learned and modify it to be more clear and comprehensive. That procedure benefits all participants. Previous studies emphasize the relationships
among important concepts and help students generate creative thoughts to make conceptual connections while doing team tasks. At the communication stage, students share knowledge and clarify misconception when they collaboratively plan and discuss the construction of CMs.

**Application.** No matter what knowledge is acquired, if the learner fails to apply it, it has little value to KT. The application of knowledge could promote familiarity of the knowledge, increase the benefits of learning and even stimulate better performance. Application can motivate students to continue the learning activity, thus improving learning quality.

Learning within practice is a matter of building up skills for students to become fully functional members of the practice community (Wenger, 1998). CM facilitates application because learners get familiar with the knowledge through practising tangible reference materials. According to Akinsanya and Williams (2004), visual learning is essential for students’ success in prioritising new information, organizing and clarifying their thinking and thus stimulating creative thinking and achieving deep learning. CMs drawn by the students show that CM provides them an organized approach for learning concepts, from the known to the unknown, from the core to the subordinate. This is an analytical approach in which knowledge is organized for cognitive retention. Kinchin, De-Leij, and Hay (2005) presented how CMs were used to integrate disparate topics and were developed into personalized mapping structures. CM has also been reported to improve students’ problem-solving ability (Okebukola, 1992).

The above characteristics of CM facilitate students’ drawing of their final CMs, which is a kind of immediate application of their knowledge. Therefore, we argue that CM could inspire learners’ application of knowledge. This knowledge can be applied in both the final presentation and future learning.

**Acceptance.** Acceptance is an intermediary process. Complete KT must go through mental acceptance. People absorb knowledge after they accept it. Thus, acceptance must occur before assimilation. Unless they are accepting the knowledge voluntarily, learners would not be satisfied with the importance of the knowledge. Acceptance is indispensable because, without acceptance, assimilation will not happen.

From the perspective of constructivist learning theory, CM is not only a self-driven learning style, but a self-directed learning experience (Teo & Gay, 2006). In other words, it makes room for learners to figure out what is suitable for them. Self-directed learning helps individuals progress in learning through multiple stages: knowledge, comprehension, application, analysis, synthesis, and evaluation (Cicognani, 2000). Moreover, in learning theory, self-regulation means the degree to which individuals are meta-cognitively, motivationally, and behaviourally active participants in their own learning processes (Zimmerman, 1986). CM fosters a productive psychological climate for learning by aiding the students’ organization of materials and focusing on crucial concepts and their features (Weinstein & Mayer, 1986). We think that CM is a learning strategy that can stimulate learners’ mental achievement. This mental achievement contributes to learning motivation. Therefore, the atmosphere encourages the acceptance of knowledge.

**Assimilation.** Assimilation is a key process in the KT framework. Setting up a formula for learning is very important. When knowledge gained is assimilated, it becomes part of the organizational culture. And then, the culture is developed into the daily routines that are reflected in the behaviour and practices of the student. In other words, the process of assimilation, or learning, occurs while knowledge and ideas are transferred within an organization to lead to the development of a set of day-to-day routines and behaviours. And this process stresses the changes from cognition to behaviour.

**The role of metacognition in KT**

Encoding is more effective when learners are aware that they are linking different items of information, rather than just storing them in isolation. Students who adopt metacognitive strategies do not only encode information in their memory, but they also may consciously look for relationships in the topics they study, such as CM (Haffker & Calvert, 2003; Scott & Schwartz, 2007). This influences their study strategies and, ultimately, how much they learn. As metacognition is further developed, learners tend to use strategies and plans to accomplish particular goals (August-Brady, 2005). As learners develop, they acquire an increasing number of strategies and they become masters whose adopted strategies can be more selective and efficient (Chularut & DeBacker, 2004).
The use of learning strategies, monitoring and reflecting, improves problem solving and increases KT when metacognition is an emphasis in the learners’ cognitive process (Eckhardt, Probst, & Schnotz, 2003). Metacognition may facilitate the integration and transfer of skills whenever confusing elements have been added to an already difficult learning situation. Metacognition, in this sense, is more likely to be problem solving (Kuhn, 2000). The learners’ involved learning activities are of great meaning when the confusing information has been transferred in an efficient way (Eggen & Kauchak, 2004). Based on the above literature, learners who adopt metacognitive strategies can employ CMs to improve problem solving and increase KT. Therefore, CM may have metacognitive traits and be related to KT.

Methodology

Participants and procedures

The sample included 42 students (33 male and 9 female) taking a nanometer course at a university in Taiwan. No one reported prior experience in CM. They were divided into 11 groups, each about four members. In the first week, each group was assigned a teaching assistant who introduced CM to the class according to the instructions of Novak and Gowin (1984). In the following weeks, the teaching assistants answered questions and provided suggestions regarding the students’ CM assignments. During the class, the instructor presented the students with both traditional lecture notes and CMs. The instructor used CMs to explain the relationship among concepts in the learning materials, as well as how to create CMs. Before the semester ended, students were asked to draw two works of CM. At the end of the semester, they were also required to fill out a questionnaire to indicate their perception of CM and their performance of nanotech KT.

The teaching assistants provided constant assistance during the period of the experiment. Meetings were held every two weeks. It was hoped that students could gain deeper understanding in order to explicitly construct a well-organized CM through the assistance of teaching assistants. In this study, based on the studies of Scheiter and Gerject (2007) and Williams (1996), students were also encouraged to ask any questions related to their task performance during the instructor’s advising hours through one-on-one contact.

Instructional design

This experiment was designed to motivate students to construct their own CMs related to knowledge of nanometer technologies (Chang, Sung, & Chen, 2001; O’Donnell, Dansereau, & Hall, 2002). Because drawing CMs alone is not easy for students who have had no prior computing drawing experience, a step-by-step instruction guide was provided. It was hoped that the learned concepts or knowledge could be illustrated on the maps:

Phase 1. Find meaningful definitions: students define professional terms.
Phase 2. Group similar sub-concepts under main concept.
Phase 3. Find links: explore relationships between two presenting concepts.
Phase 4. Make connections among joint nodes: depict relationships between two presenting concepts.
Phase 5. Illustrate examples related to learned concepts.

Procedure

Figure 2 depicts this experiment’s procedure. Students were expected to acquire the relative knowledge and skills during training.

1. Training in using FreeMind. Because students had no experience using the computer application FreeMind, training was held one week prior to the experiment.
2. Introduction to construction of integrated contents and themes in CMs via AO/GO. Students were introduced to how a well-designed CM can be structured, integrated, and organized via AO/GO. Students were encouraged to ask any questions related to their performance of tasks.
3. Begin instruction of nanometer technology.
4. Draw and submit first CMs. This was the first attempt by students to use FreeMind as a tool to delineate the relative concepts onto maps.
5. Receive comments on submissions. Students received comments and feedback on their first completed tasks.
6. Draw and submit second CMs. This was the second attempt by students in drawing the maps.
7. Receive comments on submissions. Students received comments and feedback on their second completed tasks.
8. Fill out the questionnaires.

![Figure 2. Procedure of the experiment](image)

**Measurement**

*The perception of CM.* The items on CM perception were adapted from Chiou (2008), which was originally designed to investigate students’ attitudes toward using CM to learn advanced accounting. We changed the word from “accounting” to “nanotech” in order to measure students’ perception of using CM to learn nanotech information. This measure has 11 questions to examine two dimensions of the perception. Questions 1 through 5 investigated whether the CM strategy improved the students’ learning performance. For example: “CM learning strategy stimulated me to learn and think independently.” Questions 6 through 10 examined students’ affective acceptance of the CM strategy. For example: “I was satisfied with using CM to learn nanotech.”

In Chiou’s study, this scale had high construct validity. Our revision of the scale, just one word, did not influence the high validity. The Cronbach’s α coefficient of the whole scale was 0.92 (see Table 1), and each dimension was 0.89 and 0.90, respectively, as shown in Table 1. Participants’ responses to each item were measured using a Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). A higher score indicates a more positive perception toward CM.

*The performance of KT.* Based on Gilbert’s definition of the five stages of KT, we constructed 25 questions for the five dimensions. To check their validity, the questions were submitted to one nanometer science expert and four KM experts. The questions were revised in accordance with the experts’ suggestions. The final version of the KT scale is presented in Appendix A.

The reliability verification results for internal consistency indicate that the Cronbach’s α value was higher than 0.90 (see Table 2), revealing sufficient reliability and a high degree of correlation among the questions. All of the items were self-reported from 1 (strongly disagree) to 7 (strongly agree). Higher scores show higher degrees of KT performance.
Additionally, the above two scales of the used questionnaires, KT and CM, were tested in a pilot study of 32 students using item analysis (shown in Tables 1 and 2) before establishing the Cronbach’s α reliability. The student responses to each item in the questionnaires were sorted from a low to a high. The participants were split into two groups based upon the sorted responses to each question item. The difference between both groups for each question was examined using an independent \( t \)-test. All the determinant values (\( t \) values) were found to be significant (\( p < 0.001 \)) so that each question could significantly distinguish the high-performing students from the low-performing students. Pearson’s production-movement correlation was then used to examine the extent of correlation between each question and the total score. All the correlation coefficients were found to be significant (\( p < 0.001 \)) so that there is sufficient consistency between each question and the whole test.

<table>
<thead>
<tr>
<th>CM perception</th>
<th>Item analysis</th>
<th>Item analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helpful tool item number</td>
<td>Contents</td>
<td>( t ) (sig.)</td>
</tr>
<tr>
<td>1</td>
<td>Help me learn nanotech knowledge</td>
<td>7.52 (.000)</td>
</tr>
<tr>
<td>2</td>
<td>Help me learn the concepts and the relationships of nanotech knowledge</td>
<td>10.31 (.000)</td>
</tr>
<tr>
<td>3</td>
<td>The use of learning strategies motivates my learning Improved learning sub-scale: Cronbach’s α = 0.89</td>
<td>6.61 (.000)</td>
</tr>
</tbody>
</table>

| Affective acceptance item number | Contents | \( t \) (sig.) | Correlation (sig.) |
| 5 | Enhance my motivation | 7.91 (.000) | .897 (.000) |
| 6 | Concept mapping is a new learning method | 6.37 (.000) | .449 (.041) |
| 7 | Easy of use of learning strategies to other subject matters | 8.27 (.000) | .758 (.000) |
| 8 | Attempt to use concept mapping to other subject matters | 9.41 (.000) | .880 (.000) |
| 9 | Satisfaction with the use of concept mapping | 6.83 (.000) | .775 (.000) |
| 10 | Enjoy using concept mapping | 7.26 (.000) | .459 (.036) |
| 11 | Adaptation to the use of concept mapping Affective sub-scale: Cronbach’s α = 0.90 The whole scale: Cronbach’s α = 0.92 | 7.31 (.000) | .950 (.000) |

<table>
<thead>
<tr>
<th>KT perception</th>
<th>Item analysis</th>
<th>Item analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition item number</td>
<td>Contents</td>
<td>( t ) (sig.)</td>
</tr>
<tr>
<td>1</td>
<td>Obtaining nanotech knowledge</td>
<td>7.23 (.000)</td>
</tr>
<tr>
<td>2</td>
<td>Participating in extracurricular activities</td>
<td>8.55 (.000)</td>
</tr>
<tr>
<td>3</td>
<td>Use of library resources</td>
<td>8.09 (.000)</td>
</tr>
<tr>
<td>4</td>
<td>Use of media information</td>
<td>5.95 (.000)</td>
</tr>
<tr>
<td>5</td>
<td>Use of network resources</td>
<td>7.28 (.000)</td>
</tr>
<tr>
<td>6</td>
<td>Acquiring the nanotech knowledge Acquisition sub-scale: Cronbach’s α=0.76</td>
<td>8.56 (.000)</td>
</tr>
</tbody>
</table>

| Communication item number | Contents | \( t \) (sig.) | Correlation (sig.) |
| 1 | Discussing with instructors | 7.86 (.000) | .752 (.000) |
| 2 | Listening to others’ opinions | 7.49 (.000) | .487 (.025) |
| 3 | Sharing nanotech knowledge with classmates | 8.92 (.000) | .773 (.000) |
| 4 | Joining online discussions Communication sub-scale: Cronbach’s α = 0.60 | 6.79 (.000) | .761 (.000) |

| Application item number | Contents | \( t \) (sig.) | Correlation (sig.) |
| 1 | Integrating nanotech knowledge | 6.85 (.000) | .627 (.002) |
### Data analysis

The data of the relationship between CM and KT were analyzed in two ways: canonical correlation analysis and multivariate analysis of variance (MANOVA).

#### Canonical correlation analysis

The use of canonical correlation analysis was appropriate for this study since it measures the overall relationship between two sets of variables. The method is mainly descriptive, and the results obtained from the analysis provide answers to the following questions: are the two sets of multiple variables related? And, if so, how strong is the relationship? A canonical redundancy index helps measure the ability of a set of independent variables to account for variation in a set of dependent variables. In this paper, performance of KT, including acquisition, communication, application, acceptance, and assimilation were used as one set, while CM perceptions were used as the other set.

#### Multivariate analysis of variance

The canonical correlation between CM perception and KT performance was used to examine how different levels of CM perception influence different stages of KT (acquisition, communication, application, acceptance, and assimilation). The data were also analyzed by MANOVA. The MANOVA analysis is appropriate when there are multiple dependent variables as well as independent variables within the model which the researcher wishes to test.

#### Research results

Canonical correlation analysis was performed to assess the relationship between CM perception (CMP) and KT performance (KTP). The analysis yielded two roots. We report only the results of the first function because the second function is not significant. The first function, $R_c = 0.81$, Wilks’s $\lambda = 0.28$, $F(10, 42) = 46.52$, $p < 0.001$, accounted for 65% of the common variance between CMP and KTP. The structure of the relationships between CMP and KTP, together with their canonical correlations, is shown in Figure 3.
A canonical redundancy index helps measure the ability of a set of left-side variables (e.g., measures of the perception of CM) to account for variation in a set of right-side variables (e.g., measures of KT performance). Table 3 summarizes the structure loading and redundancy index analyses for both the predictor set and the criterion set in the study. The results indicate that the predictor set explained 53.3 percent of the variance in the criterion set. As the canonical loadings reveal, both improved learning (0.93) and affective acceptance (0.88) loaded high on CM perception. They exhibited their strong links to the five stages of KT. Among the five stages, acceptance emerged as the most important stage (0.93), and assimilation is the second most important one (0.84).

The function indicated that improved learning and affective acceptance of CM significantly and positively correlate with the five stages of KT: acquisition, communication, application, acceptance, and assimilation. The practical meaning of the result is that if students have a more positive perception toward CM, they are more likely to perform KT.

To further explore the relationships among the five stages of KT and students’ perception of CM, MANOVA was employed to examine how different levels of CM perception influence different stages of KT (acquisition, communication, application, acceptance, and assimilation). First, CM perception was transformed into a categorical variable with three values: low, medium, and high. The top 26% and bottom 26% were set as thresholds, as described in Table 4. The five stages of KT were used as dependent variables. The result of MANOVA analysis showed whether different levels of CM perceptions produced significantly different performance outcomes of KT in terms of the five stages.

<table>
<thead>
<tr>
<th>CM perception label</th>
<th>Range</th>
<th>Mean (SD)</th>
<th>N</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster1—low perception</td>
<td>2.73 – 4.45</td>
<td>3.91 (0.61)</td>
<td>11</td>
<td>Bottom 26%</td>
</tr>
<tr>
<td>Cluster2—medium perception</td>
<td>4.64 – 5.64</td>
<td>5.11 (0.32)</td>
<td>20</td>
<td>48%</td>
</tr>
<tr>
<td>Cluster3—high perception</td>
<td>5.73 – 7</td>
<td>6.26 (0.40)</td>
<td>11</td>
<td>Top 26%</td>
</tr>
<tr>
<td>Total</td>
<td>2.73 – 7</td>
<td>5.10 (0.96)</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

The results of the MANOVA analysis were shown in Table 5. Multivariate tests revealed that the overall model were significant (Wilks’s $\lambda = 0.26$, $p < 0.001$), indicating that CMP has a significant effect on KTP.
Table 5. Test index of MANOVA analysis

<table>
<thead>
<tr>
<th>Effect</th>
<th>Test index</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMP</td>
<td>Pillai’s trace</td>
<td>0.86</td>
<td>5.44</td>
<td>10</td>
<td>72</td>
<td>0.000</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Wilks’s lambda</td>
<td>0.26</td>
<td>6.62</td>
<td>10</td>
<td>70</td>
<td>0.000</td>
<td>0.48</td>
</tr>
<tr>
<td>Level</td>
<td>Hotelling’s trace</td>
<td>2.31</td>
<td>7.86</td>
<td>10</td>
<td>68</td>
<td>0.000</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Roy’s largest root</td>
<td>2.08</td>
<td>15.00</td>
<td>10</td>
<td>36</td>
<td>0.000</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Furthermore, individual ANOVAs were computed on the scores of the five KT stages. The results in Table 6 show significant differences among the three clusters at each KT stage: acquisition, $F(2, 39) = 8.71, p < 0.001$, communication, $F(2, 39) = 11.59, p < 0.001$, application, $F(2, 39) = 20.71, p < 0.001$, acceptance, $F(2, 39) = 29.03, p < 0.001$, and assimilation, $F(2, 39) = 9.39, p < 0.001$. In summary, different levels of CM perception had significantly different effects on performance outcomes of KT in the five KT stages.

Table 6. Summary of one-way ANOVA and post hoc test

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Dependent variables</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Post hoc tests (Scheffe tests)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM perception</td>
<td>Acquisition</td>
<td>6.22</td>
<td>2</td>
<td>3.11</td>
<td>8.71***</td>
<td>3 &gt; 1, 2 &gt; 1</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>8.47</td>
<td>2</td>
<td>4.24</td>
<td>11.59***</td>
<td>3 &gt; 1, 2 &gt; 1</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>20.24</td>
<td>2</td>
<td>10.12</td>
<td>20.71***</td>
<td>3 &gt; 1, 3 &gt; 2</td>
</tr>
<tr>
<td></td>
<td>Acceptance</td>
<td>13.83</td>
<td>2</td>
<td>6.92</td>
<td>29.03***</td>
<td>3 &gt; 2 &gt; 1</td>
</tr>
<tr>
<td></td>
<td>Assimilation</td>
<td>11.01</td>
<td>2</td>
<td>5.50</td>
<td>9.39***</td>
<td>3 &gt; 1, 3 &gt; 2</td>
</tr>
</tbody>
</table>

*** p value < 0.001

As described in Table 7, because all $F$-values were significant, a series of post hoc tests (Scheffe’s tests) were also conducted to further compare the three clusters. The results revealed that the students in cluster three had a significantly higher achievement than those in cluster one. That is to say, students in cluster three performed more “acquisition” (5.86 versus 4.8, $p < 0.001$), and “communication” (5.77 versus 4.57, $p < 0.001$) than those in cluster one. Additionally, their performance in “application” (6.29 versus 4.49, $p < 0.001$), “acceptance” (6.27 versus 4.71, $p < 0.001$) and “assimilation” (5.91 versus 4.56, $p < 0.001$) was better than students in cluster one and cluster two. The detailed information is listed in Table 7. The results of Scheffe’s tests reveal that significant differences were found between cluster one and cluster three across all KT stages. Consequently, we concluded that students with higher CM perception tended to perform better KT than those with lower perception.

Table 7. KT of the three clusters

<table>
<thead>
<tr>
<th></th>
<th>Acquisition mean (SD)</th>
<th>Communication mean (SD)</th>
<th>Application mean (SD)</th>
<th>Acceptance mean (SD)</th>
<th>Assimilation mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>4.80 (0.69)</td>
<td>4.57 (0.50)</td>
<td>4.49 (0.54)</td>
<td>4.71 (0.55)</td>
<td>4.56 (0.99)</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>5.39 (0.55)</td>
<td>5.39 (0.64)</td>
<td>4.91 (0.87)</td>
<td>5.30 (0.51)</td>
<td>4.92 (0.67)</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>5.86 (0.58)</td>
<td>5.77 (0.64)</td>
<td>6.29 (0.43)</td>
<td>6.27 (0.35)</td>
<td>5.91 (0.65)</td>
</tr>
<tr>
<td>Total</td>
<td>5.36 (0.70)</td>
<td>5.27 (0.74)</td>
<td>5.16 (0.98)</td>
<td>5.40 (0.75)</td>
<td>5.09 (0.91)</td>
</tr>
</tbody>
</table>

Discussion and conclusions

The study investigated how students used CM integrated via AO/GO as a learning strategy and the progress of KT. We examined the learners’ attitudes toward CM and investigated whether their perceptions caused differences in their learning outcomes. We also presented the characteristics of each KT stage, illustrating knowledge conversion for operation. The data analysis here also provided empirical evidence for the meaningful study in nanotech of science domain.

The results of canonical correlation showed that students were in favor of using CM and this attitude significantly and positively influences their KT performance. That is, if students express higher perception toward CM, they are more likely to perform KT. The results reveal that significant differences were found between the students who have
highly positive perception of CM and those who have low perception, in every stage of KT: acquisition, communication, application, acceptance, and assimilation. In summary, students with higher perceptions perform more KT than those with lower perceptions.

Previous research mainly investigated the relationship between CM and learning achievement. However, this study broadens the application of CM to individual KM focusing on KT performance. Our findings are similar to those from previous studies in other disciplines, demonstrating that the CM strategy had positive effects on student performance (Chiu, 2008; Kinchin, 2001; Kinchin, De-Leij & Hay, 2005; Lawless, Smee, & O’Shea, 1998). Most of these studies showed that CM brings high and useful effects while learning outcome assessment. For example, Chiu (2008) revealed that CM is useful in fostering students to improve their learning performance in a business and economics statistics course. CM is better than traditional textbook exercises. Most of the students were satisfied with using CM to learn. On the other hand, Lawless, Smee, and O’Shea (1998) mentioned that the use of CMs in business and public administration is concentrated on group decision-making and planning. Through the process of constructing CMs, various viewpoints can be bracketed into categories, which are refined into an index and taken as a guideline for actions. This procedure not only has positive effects on the coordination of different perspectives across institutions and companies but also has benefits for action plans.

In this study, the CM strategy was applied to KM. The results illustrate that there was a strong connection between CM and KT. The findings, consistent with our proposition, indicated that students understand the advantages of the CM and are willing to adopt it because it stimulates the performance of KT. To achieve KT, teachers’ materials should use AO/OI to better organize and integrate knowledge in order to let CM meet the needs of students, considering that it can improve students’ learning perception, accumulation of knowledge, sharing of information, and KT. In addition, previous studies mentioned that students who adopt metacognitive strategies may consciously look for relationships in CM (Haffker & Calvert, 2003; Scott & Schwartz, 2007) and can use these strategies to improve problem solving and increases KT (Eckhardt, Probst, & Schnotz, 2003). Therefore, students may employ CMs when metacognition, the awareness and ability to monitor and reflect upon knowledge, is an emphasis in their cognitive learning process. Metacognition may also facilitate the knowledge integration and transfer. However, individual differences in the level of metacognitive skill, ability, and prior knowledge, etcetera, between students may play a major role if students are low-ability or low prior knowledge learners. Such individual differences may impact the significance of our findings.

Since CM is new to these participants, the students need some effort and time to become accustomed to it. Their perception toward CM influenced learning effects and further affected the consequence of KT. Therefore, it is important to make sure that students possess the capability of using CMs and have positive attitudes toward the measurement before assessment. The results from this study found that there was a significant relationship between CM perception and KT performance outcome. The finding is similar to results of past studies (González et al., 2008; Mann & LeClair, 2009; Pinto, 1997) showing that perceived CM had been associated with KT performance outcome in different disciplines. The researchers conclude that perceived CM promotes the transition among the five stages of KT performance outcome. The limitation of this study was a concern related to the sample selection. Because the experimental class was the only convenient class that could be used at the time the experiment was conducted, bias of sample selection may be a threat to the internal validity in this study.

The result of the study showed that positive perception increases the willingness to use CM. It is argued that through improving the learning attitudes of the students, the instructor can get students to be more satisfied with their learning and thereby make the KT process more effective. That progress encouraged students to produce meaningful learning through CM. It allowed students to organize and share information efficiently through CM. It was beneficial for individual’s curriculum learning and KM.

The core idea of this research was to investigate the use of CMs as a technique that smoothes the transition among the five stages of KT: acquisition, communication, application, acceptance, and assimilation. We conclude that the technique of employing CMs is an appropriate tool to promote students’ positive perception of learning and to facilitate the transitions among these stages of KT. This study confirmed the link between CM perceptions and KT with empirical evidence.
Acknowledgements

Funding for this research was supported by National Science Council, Taiwan, and the principal investigator was Dr. Kuo-hung Tseng. The author appreciates the kind assistance of Dr. Ling-Ling Yen, Miss Ting-Hsiu Liu, and also the helpful comments from the reviewer, who made this paper possible.

References


Appendix A: Survey of self-evaluation of knowledge transfer

Stage 1: Acquisition
I obtain nanotech knowledge by class learning such as in a school, a cram school, etc. I participate in extracurricular activities (such as listening to speeches, visiting with friends, participating in competitions) to broaden the horizons of nanotech knowledge. I utilize library resources (for instance, books, journals, online research resources) to increase my nanotech knowledge. I use media information (for instance, televisions, broadcasts, newspapers) to enrich my nanotech knowledge, I use network resources (for example, search engines, blogs, Yahoo! Answers) to expand my nanotech knowledge. When I meet problems, I know where to acquire the nanotech knowledge.

Stage 2: Communication
Discussing with instructors for the nanotech knowledge clarifies my concepts. Listening to others’ opinions improves the understanding of nanotech knowledge. Sharing nanotech knowledge with classmates deepens my knowledge of nanotech. Joining online discussions (such as bulletin boards, web logs, Yahoo! Answers) reinforces my realization of nanotech knowledge.

Stage 3: Application
I can integrate nanotech knowledge with my original knowledge. I can draw concept maps with my learned nanotech knowledge. I can demonstrate the concept mapping with my learned nanotech knowledge. The accumulation of the experience of nanotech learning improves my appliance of the knowledge. Knowledge and understanding of nanotech inspires me to learn more in related areas.

Stage 4: Acceptance
Experiencing nanotech knowledge raises the degree of accepting it. My understanding of nanotech knowledge extends my identification on its value. My understanding of the nanotech knowledge lifts my acceptance of it. In-depth exploration of nanotech knowledge motivates me to keep learning. After gaining the nanotech knowledge, I am satisfied with the whole learning content.

Stage 5: Assimilation
I will actively teach others my learned nanotech knowledge. I will apply my learned nanotech knowledge on the present subjects. I will apply my learned nanotech knowledge on the future relevant subjects. I will apply my learned nanotech knowledge on solving the life problems. I will expand my learned nanotech knowledge on future projects.
Elementary School Students’ Perceptions of the New Science and Technology Curriculum by Gender

Mehmet Nuri Gömleksiz
Department of Educational Sciences, Faculty of Education, Fırat University, Elazig, Turkey // nurigomleksiz@yahoo.com

ABSTRACT
The purpose of this study is to explore students’ perceptions of science and technology classes by gender in a Turkish elementary school context. Data for the study were collected through a 20-item, five-point Likert scale from a total of 1558 sixth-grade students at 20 different elementary schools in Turkey. The independent groups’ t-test and Mann-Whitney U test were used to analyze the data. Statistically significant differences were observed in the gender of the students. Male students considered learning science and technology more necessary and important than female students did. They also found learning environment and teaching strategies more sufficient and effective than females did. Findings revealed that male students were not satisfied with what the teachers practised in science classrooms. Additionally, some useful implications are discussed based on the research findings to construct and conduct science and technology classes effectively.

Keywords
Elementary schools, gender, science teaching, science curriculum, Science education, technology education

Theoretical framework and background of study
In Turkey, curriculum development activities started with the foundation of the Modern Turkish Republic in 1923. Reforms of many curricula have been developed and implemented at schools so far (Gozutok, 2003; Basar, 2004; Babadogan & Olkun, 2006). Turkey has always made major reforms in the area of curriculum development at the elementary school level to improve the quality of education, and a new elementary school curriculum, including science and technology, has been completely changed and implemented nationwide starting with the 2005–2006 academic year. These changes included both the name and the content of the science courses, and “science education” was changed to “science education and technology” (Turkmen, 2006).

The aim of the new science and technology curriculum is to provide a student-centered learning environment based on a cognitive and constructivist learning approach instead of a rigid and strict behavioral approach. The principles of multiple intelligences and active learning based on individual differences have also been adopted with the new science and technology curriculum. Students are expected to gain the following skills that they previously lacked: critical and creative thinking, communication, scientific research, problem solving, using information technologies, and entrepreneurship. Students are also expected to become science and technology literate with the new science and technology curriculum. They are required to understand the basic concepts of science and technology and to relate technological and scientific knowledge to each other and to the world outside the school. The increasingly complex changes in the nature and amount of knowledge and demands in the field of science and technology necessitate an understanding of how students perceive science and technology classes in terms of their gender. Differences in perceptions of science and technology between boys and girls have been examined by many scholars (Kahle, 1983; Raat & de Vries, 1985; Baker, 1987; Collis & Williams, 1987; Kurth, 1987; Piburn & Baker, 1989; Bame, Dugger, de Vries & McBee, 1993; Weinburgh, 1995; Speering & Rennie, 1996; Baker, 1998; Francis & Greer, 1999; Udo, Ramsey, & Mallow, 2005; Ogunjugibe, Ojoefitimi, & Akinlo, 2006). The studies have reported that male students have greater interest and achievement than female students in science and technology. Specifically, Boser, Palmer, and Daugherty, (1998) reported that female students consistently perceived technology to be less interesting than male students did. In other related studies, Jewett (1996) and Silverman and Pritchard (1993) found that technology, mathematics, and science are still considered nontraditional areas for females and that some societal perceptions and expectations contribute to women’s reduced interest in these fields. In fact, the most striking difference between males and females in science is not in achievement or in opportunities to learn, but in confidence. Even when females have similar exposure to courses and a similar achievement level, they are less confident in their ability, feel less prepared, and lack interest in science and technology (Lundeberg, Fox, & Puncochar, 1994; Sax, 1995; Seymour & Hewitt, 1997). The results of several studies show that the overall trend for male students’ perceptions about the utility, necessity, and importance of science and technology is positive (Kahle & Lakes, 1983; Erickson & Erickson, 1984; Johnson, 1987; Meyer & Koehler, 1990; Erickson & Farkas, 1991; Greenfield, 1997; Jovanovich & King, 1998; Spall, Barrett, Stanistreet, Dickson & Boyles, 2003). This means that...
male students have positive perceptions about science and technology classes. However, there are some other studies indicating that female students perceive the biology learning environment more favorably than male students do (Waxman & Huang; 1998; Dawson, 2000; Jones, Howe & Rua, 2000; Prokop, Tuncer & Chuda, 2007). Popham (1994) suggests that even affective behaviors are acceptable to undergo far more sudden transformations than cognitive behaviors. It is also possible that if students have a tendency to act positively toward a subject, for example, science and technology, then students will have a greater interest in those subjects (Krathwohl, Bloom, & Bertram, 1964).

The studies have often investigated associations between student outcomes and the nature of the classroom environment and showed that the conditions of learning environment affect students’ beliefs and success in science and technology classes (Hofstein, Scherz, & Yager, 1986; Talton & Simpson, 1987). Educational environments enhance students’ learning and improve academic achievement (Massachusetts Department of Education, 2006). A well-designed learning environment aimed at providing effective instruction enriches learning experiences as well. Students should be aware of what they really need and what they should know. Just as “learning environment” refers to the factors that can affect a person’s learning, “social environment,” which includes family members and friends in a wider context, affects the learner and shapes his/her learning. Therefore, students should be provided a rich and supportive learning environment. Effective teaching requires a combination of many factors, including aspects of the teacher’s background, and ways of interacting with others, as well as specific teaching practices. Effective teachers care about their students and demonstrate this care in such a way that their students are aware of it. This care creates a warm and supportive classroom environment (Stronge, 2002). Teachers have a profound effect on student learning. They can bring the real world to students through technology and can facilitate teaching (Schroder, Scott, Tolson, Huang, & Lee, 2007).

One of the primary reasons students fail in science is because they often have learning styles significantly different from those emphasized by most science courses (Felder, 1993). As individuals have different preferences in giving meanings and acquiring information, the ways in taking and processing information may vary (Yılmaz-Soylu & Akkoyunlu, 2009). While some prefer to work with concrete information, others are more comfortable with abstractions. Some learn better by visual presentations such as diagrams, flowcharts, and schematics; others learn more from verbal explanations (Felder & Spurlin, 2005). If students’ learning styles match the teaching style of the teacher, students will keep information longer and apply it more effectively (Felder, 1993).

Cooperative learning is one of the teaching strategies used effectively in science and technology classes. In previous studies, females rated cooperative school activities more positively than did males (Shwalb, Shwalb, & Nakazawa, 1995; Ferreira, 2003). Owens and Stratton (1980) found that girls prefer cooperation, open-ended, and organized activities, while boys prefer competition and individualism. By using cooperative learning practices, learning is maximized and both positive and productive interactions are provided between students of different backgrounds (Cabrera, Crissman, Bernal, Nora, & Pascarella, 2002). All these studies prove the importance of designing effective learning environments and using teaching strategies that will enhance students’ improvement in science and technology classes. Creating a student-centered, creative, and effective learning environment allows students to express themselves better and gives them the opportunity to understand themselves in terms of strengths and weaknesses when they study science and technology.

In brief, the educational contexts or curricular programs in which elementary school students enroll play an important role in their perceptions of learning science and technology. With the current study, it was aimed to determine how sixth-grade elementary school students perceived science and technology courses and whether there were any differences between their perceptions based on gender. This study outlines a framework to describe the variations of the perceptions of learning science and technology, consisting of the following features: the need to see science and technology as necessary and important, the learning environment involved, and teaching strategies employed. How to construct and conduct science and technology courses effectively and sufficiently at the elementary school level was also discussed.

Method

Purpose of study

The present study was an attempt to explore sixth-grade elementary school students’ opinions about science and technology courses implemented at 20 elementary schools in Elazig, Turkey. We aimed to see whether there were
any statistically significant differences among elementary students’ views toward importance, necessity, learning environment of science and technology classes and teaching strategies of their science teachers in terms of the gender variable.

Population and sampling

The population of this study comprised sixth-grade students from 20 elementary schools in Elazig, Turkey. The stratified proportional random sampling procedure was used to select the schools. There are five education zones in the city. The schools were stratified on the basis of education zones and their socioeconomic conditions. Then four schools from each education zone were selected with three levels of socioeconomic status reported by the National Education Office. The sample consisted of 1,558 (925 male and 633 female) sixth-grade students selected randomly from those elementary schools. The simple random sampling procedure was used to select 318 students from each elementary school. A total of 1,590 students participated in the study. However, out of these 1,590 students, 1,582 completed questionnaires. Of this total number, 24 were incomplete and were thus eliminated, leaving a sample of 1,558 students. The gender composition of the respondents is 59.4% male and 40.6% female.

Data collection and analysis

Data was generated from a questionnaire in which a five-point Likert scale was used. Students were asked to rate their opinions about the science and technology courses they were taking (Appendix A). The survey, administered in the classroom, included 20 items derived from the review of the relevant literature. In some classes the researcher administered the survey himself and in others, classroom teachers were trained to administer the survey. In all cases, the same procedures were followed. Students were reminded that their answers would remain anonymous, and they were asked to read the items carefully and answer honestly.

The survey, with responses of strongly agree, agree, partly agree, disagree, and strongly disagree, was first piloted on 415 students for the factor-analysis process. The pilot participants were similar to the target population in terms of background. Factor analysis was used to examine the correlation between the items of each subscale. Factor analysis revealed four subscales, namely: importance, necessity, learning environment, and teaching strategy.

<table>
<thead>
<tr>
<th>Subscales</th>
<th>Item No</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>1, 3, 7, 11, 18</td>
<td>0.94</td>
</tr>
<tr>
<td>Necessity</td>
<td>2, 9, 10, 13, 14</td>
<td>0.94</td>
</tr>
<tr>
<td>Learning environment</td>
<td>4, 5, 12, 15, 17</td>
<td>0.93</td>
</tr>
<tr>
<td>Teaching strategy</td>
<td>6, 8, 16, 19, 20</td>
<td>0.94</td>
</tr>
</tbody>
</table>

The internal reliability of the scale was calculated by using Cronbach’s alpha formula, the Spearman-Brown reliability coefficient, and Guttmann’s split-half technique. Cronbach’s alpha for the importance subscale ($\alpha = 0.94$), necessity subscale ($\alpha = 0.94$), learning-environment subscale ($\alpha = 0.93$), teaching-strategy subscale ($\alpha = 0.94$), and overall scale ($\alpha = 0.98$) showed satisfactory reliability because a Cronbach’s alpha scale greater than 0.70 was considered acceptable for the internal reliability of the items associated with each proposed factor (Hair, Anderson, Tatham, & Black, 1995). For the whole sample, the Spearman-Brown reliability coefficient for unequal lengths was calculated to be 0.96, and Guttmann’s split-half technique revealed a reliability coefficient of 0.96. The Kaiser-MeyerOlkin measure of sampling adequacy of the scale was measured to be 0.97, and Bartlett’s test was calculated to be 37,063.668 ($p < 0.05$). According to the results obtained from the factor-analysis process, the scale was found to be valid and reliable. In a prior examination, when the distribution of the data was found to be non-normal, the non-parametric statistical technique Mann-Whitney U was used for testing gender differences. When the distribution of the data was found normal, the parametric statistical technique, the independent groups $t$-test, was used. Results with $p < 0.05$ were considered statistically significant.
Results

The first analysis was to determine if any significant differences between the students’ views existed on the “importance” subscale.

Table 2. t-test results for gender on the “importance” subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Gender</th>
<th>n</th>
<th>$\bar{X}$</th>
<th>sd</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>Male</td>
<td>925</td>
<td>3.91</td>
<td>0.81</td>
<td>1556</td>
<td>6.029*</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>633</td>
<td>3.66</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.05 level

As shown in Table 2, statistically significant differences were found in terms of gender of the students [$t_{(1556)} = 6.029$, $p < 0.05$]. A higher mean rating suggested that male students were more in agreement with the importance of science and technology courses than the female students were.

Table 3. Mann-Whitney U results for gender on the “necessity” subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Gender</th>
<th>n</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
<th>MWU</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necessity</td>
<td>Male</td>
<td>925</td>
<td>831.42</td>
<td>769068.00</td>
<td>244732.0*</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>633</td>
<td>703.62</td>
<td>445393.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.05 level; Levene: 10.802, $p < 0.05$

Table 3 presents the summary of analysis Mann-Whitney U comparing the mean scores of the male and female students in terms of necessity of the science and technology classes. With regard to gender differences, it appears from the data that there was significant gender difference on the necessity to learn science and technology (MWU = 244732.0, $p < 0.05$). The statistically significant difference between gender groups suggests that male students had higher mean scores than did female students. Male participants accepted the necessity of science and technology more than female students did.

Table 4. Mann-Whitney U results for gender on the “learning environment” subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Gender</th>
<th>n</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
<th>MWU</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>Male</td>
<td>925</td>
<td>834.21</td>
<td>771644.50</td>
<td>242155.5*</td>
<td>0.000</td>
</tr>
<tr>
<td>environment</td>
<td>Female</td>
<td>633</td>
<td>699.55</td>
<td>442816.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.05 level; Levene: 4.925, $p < 0.05$

Mann-Whitney U results in Table 3 revealed statistically significant differences between the student groups in gender (MWU = 242155.5, $p < 0.05$). The significant MWU value obtained for gender demonstrated that female students found the learning environment in science and technology classes less sufficient and effective than did male students.

Table 5. t-test results for gender on the “teaching strategy” subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Gender</th>
<th>n</th>
<th>$\bar{X}$</th>
<th>sd</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>Male</td>
<td>925</td>
<td>3.88</td>
<td>0.80</td>
<td>1556</td>
<td>5.919*</td>
<td>0.000</td>
</tr>
<tr>
<td>strategy</td>
<td>Female</td>
<td>633</td>
<td>3.64</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at the 0.05 level

As illustrated in Table 5, t-test results revealed significant differences between male and female students in terms of teaching strategy of the science and technology teachers [$t_{(1556)} = 5.919$, $p < 0.05$]. Female participants tended to adopt teaching strategies used in science and technology classes less efficiently and effectively than did male students.

Discussion

The quality of learning has always been one of the most important concerns in an educational setting. Learning is a complex activity, and several factors such as students’ perceptions, beliefs, and attitudes; teaching resources; teachers’ skill; curriculum; physical condition; and the design of the school facility should be taken into
consideration in an educational setting. They all play a vital role in providing effective education (Lyons, 2001). The quality of learning experience can be understood through an investigation of how key factors of the experience are related. Key factors associated with the quality of the learning experience are how students approach their learning and what they think they actually learn from the experience (Ellis, 2004).

The present research evaluated and compared sixth-grade students’ perceptions of science and technology classes at 20 different elementary schools in Turkey. The results of the present study show that sixth graders’ perceptions of science and technology classes differed significantly by gender. One of the most significant conclusions to be drawn from the findings was that male students were interested in science and technology classes more than female students were. The result derived from the findings of the current study is consistent with results from previous research (Erickson & Erickson, 1984; Johnson & Murphy, 1984; Simpson & Oliver, 1985; Johnson, 1987; Becker, 1989; Engstrom & Noonan, 1990; Greenfield, 1996; Lee & Burkam, 1996; Ding & Harskamp, 2006).

Statistically significant differences were found between male and female students toward the importance of science and technology classes. It means that the data supports the significance difference between male and female students’ perceptions toward the importance of science and technology. Male students considered science and technology classes more important and had a more positive tendency toward learning science and technology than the female students did. Gender issues have long been a topic of discussion and research in the field of science and technology education. Numerous studies have been conducted to explain gender differences in participation and achievement in science and technology. Studies show that many instructors base their expectations of student performance on gender factor as well as language proficiency, socioeconomic status, and prior achievement (Green, 1989). Leder (1989) in particular has claimed that academic success in mathematics is associated mostly with males. The results of other studies clearly showed that male students consistently showed a higher interest and achievement than females (Johnson, 1987, Tobin & Garnett, 1987; Norman, 1988; Otto 1991; Meece & Holt 1993; Trumper, 2006). Taking these results into consideration, many science education programs have recently been developed to increase girls’ participation in science (Yanowitz & Vanderpool, 2004).

There was statistically significant difference between gender groups toward the necessity of science and technology classes. This finding indicates that male students found science and technology classes more necessary than female students did. This result is consistent with the findings by James & Smith (1985), Eccles (1989), Linn & Hyde (1989), Kahle & Meece (1994), and Catsambis (1995), who found that gender differences begin to appear in the middle grades and that the gender gap in science achievement increases between ages 9 and 13. This result also supports Yager & Yager (1985), Schibeci (1984), Greenfield (1996), Jovanovich & King (1998), and Stake & Mares (2001), who found that students begin to show differences for science in elementary and middle school, and that girls exhibit lower science achievement scores than boys do at the middle-school level.

Statistically significant differences were found between gender groups toward their perceptions of learning environment in science and technology classes. Female students found the learning environment less sufficient and effective than did male students. The quality of the learning environment is important to realize effective learning, and a well-designed learning environment both enhances students’ learning and leads to higher learning achievement. It not only depends on the design but also on how effectively it is delivered and used because the learning experience is directly influenced by the way the learning resource is delivered. To do this, the learning environment should be designed to promote relevant interaction between learner and learning resources to achieve the stated learning outcomes and to provide timely feedback to learners regarding their progress, and should be consistent with the most efficient and effective method to meet learning outcomes.

The findings of the present study imply that it is important to design learning environments in such a way as to facilitate and enhance science and technology learning. These are in line with the ideas of contemporary learning theorists such as Brown, Collins, and Duguid (1989); Spiro, Feltovich, Jacobsen, and Coulson, (1991); and Bereiter and Scardamalia (1996). They believe that one of the key goals of instruction is to provide opportunities for learners to develop mastery in the areas they are each involved in. School facilities have also an effect on student performance. Recent studies that evaluated the relationship between school buildings and student achievement found higher test scores for students learning in better buildings and lower scores for students learning in substandard buildings. A recent report evaluating school facilities showed a difference in student test scores ranging from 5 to 17 percentage points (Lyons, 2001).
A significant difference was found between the gender groups in terms of teaching strategy used in science and technology classes. Male students found teaching strategies more effective and sufficient while female students found them insufficient. Teaching strategy used in the classroom has a direct influence on how a teacher manages the classroom. Teachers must design teaching and learning strategies around students’ interest to improve the quality of the learning environment. For instance, the use of inquiry-based approaches in a science classroom leads students to understand the way science is authentically carried out. Many studies have proved that inquiry-based science activities have positive effects on student achievement, cognitive development, laboratory skills, and the understanding of science content when compared with traditional approaches (Burkam, Lee, & Smerdon, 1997; Freedman, 1997).

Effective use of teaching strategies encourages students in a positive and supportive manner and helps them participate actively in the teaching-learning environment. Both a growing student population and student diversity require changes in how students are taught. As Labudde, Herzog, Neuenschwander, Violi, & Gerber (2000) stressed, strategies should include opportunities to integrate different pre-existing knowledge and the variation of teaching methods to enhance cooperation and communication in the classroom. Because each student learns in different way and has his/her own learning style, an approach that is appropriate for one student may be inappropriate for another. While some students learn better in a group through interaction with both the teacher and other students, others may find interaction difficult and use the group sessions for gathering information. They learn only when they are on their own. Some learn by reading and listening, while others learn through the application of the knowledge gained. Teachers should concentrate on such differences and enrich the learning environment by providing a variety of learning activities so that students can learn in a manner appropriate to themselves (Reece & Walker, 1997). Therefore, it is vital that teachers guide their students to actively participate in the learning environment.

**Conclusion and recommendations**

The findings of this study are subject to two limitations. First, the data apply only to the 1,558 sixth graders who attended 20 different elementary schools in Turkey. Second, the findings cannot be generalized to evaluate the overall effectiveness of the science and technology classes in elementary schools throughout Turkey. This is not because this particular region in Turkey is extremely different from other regions. The particular research region was chosen because the researcher works there. Because the sample was selected by stratified proportional random sampling procedure, it represents the city that was investigated. Despite limited generalizability, this study represents an attempt to understand student perceptions of their science and technology classes in terms of the gender variable. The results from this study identified areas of strengths and weaknesses within elementary schools’ science and technology classes from students’ perspectives.

Science teachers should concentrate on authentic activities. The learning subject taught in the activity must suit students’ ages, interests, expectations, and environment. Students must be able to use what they learned in science classes. Teachers should not be dependent on the textbooks strictly to provide a more flexible learning environment. They should sometimes feel free to adapt textbook activities and avoid mechanical activities. To achieve this, diversity of activities is needed. The activities should be performed in pair or group work so that the students can build a cooperative learning environment. Research results have shown that cooperative activities facilitated more active roles and enhanced students’ learning (Baker, 1990; Meyer, 1998; Bilgin & Geban, 2004, Açıkgöz & Güvenç, 2007). Students learn better in a group through interaction with both the teacher and other students. Grouping for cooperative learning activities based on gender may lead students to learn better and promote positive attitudes. Mixed-gender groups in particular show better achievement and improvement. So, the different learning and motivational styles of males and females should be taken into consideration (Kemp, 2005). Research studies have confirmed that females focus on completing a task correctly whereas males are often more motivated to be better than everyone else at completing a task (Rogers et al., as cited in Kemp, 2005). Females may have a fear of making mistake under the pressure of a difficult task and may withdraw from the activity. But performing a difficult task may motivate males (Dai, 2000).

In determining the design of the learning environment, the importance of a variety of learning activities for students should be taken into consideration. The teachers should implement learning strategies that will encourage female students to engage in science and technology classes and to narrow the gap between male and female students for participating in teaching-learning activities. Previous studies showed that psychosocial climate and physical
conditions of a learning environment have an important effect on students’ outcomes (Fraser, Williamson, & Tobin, 1987; Lawrenz, 1987; Talton & Simpson, 1987; Schibeci, Rideng & Fraser, 1987; Fraser, 1998; Panagiotopoulou, Christoulas, Papanckolaou, & Mandroukas, 2004). The teaching-learning activities in science and technology classes should be purposeful and meaningful. Students should be given convincing reasons for doing the activity, and they should know what they would have achieved upon completion of the activity. Students should not only be physically active but also mentally active in the learning process (Babadogan & Olkun, 2006).

The findings of this research reveal that science and technology teaching in Turkish elementary schools needs a radical overhaul to attract students’ interest and increase participation in science and technology classes. The results of the current study confirm the earlier findings that there are gender differences in science and technology achievement. Understanding some of the concerns of elementary school students with regard to science and technology teaching might help curriculum designers and teachers, as practitioners, modify or change existing programs to meet the requirements of the students and of the content area. Although the results of this study provide information about the perceptions of students on science and technology classes, additional research is needed to better understand how the science and technology curriculum is implemented and whether it is conducted effectively and sufficiently.

References


Appendix A

Science and Technology Scale

Students were requested to respond to the following statements on a five-point Likert scale ranging from strongly agree through to strongly disagree.

1. The science and technology course is important.
2. The knowledge I gain in science and technology class is always helpful in real life.
3. Science and technology affect our life very much.
4. I feel confident in science and technology class.
5. The learning environment of science and technology class encourages me to participate actively.
6. I find the teaching strategy used in science and technology class effective.
7. The science and technology course includes important knowledge that I may need in the future.
8. The teaching strategy used in science and technology class makes me attentive to the lesson.
9. I will always need the knowledge gained in science and technology class.
10. I need to become science and technology literate.
11. We need to be taught science and technology to prepare ourselves for the future.
12. I find the learning environment in science and technology class effective.
13. Science and technology is a subject that should be taken by all pupils.
14. Science and technology courses are necessary to help us develop our creativity.
15. We are encouraged to research and study in science and technology courses.
16. The teaching strategy of the science and technology teacher encourages me to be active during class.
17. The learning environment of the science and technology course is both supportive and collaborative.
18. Science and technology courses give us the opportunity to learn by doing.
19. I like my science and technology teacher’s teaching strategy (methods).
20. My science and technology teacher can teach clearly and comprehensibly.
Student Satisfaction, Performance, and Knowledge Construction in Online Collaborative Learning

Chang Zhu
Department of Educational Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium // chang.zhu@vub.ac.be

ABSTRACT

A growing amount of research focuses on learning in group settings and more specifically on learning in computer-supported collaborative learning (CSCL) settings. Studies on western students indicate that online collaboration enhances student learning achievement; however, few empirical studies have examined student satisfaction, performance, and knowledge construction through online collaboration from a cross-cultural perspective. This study examines satisfaction, performance, and knowledge construction via online group discussions of students in two different cultural contexts. Students were both first-year university students majoring in educational sciences at a Flemish university and a Chinese university. Differences and similarities of the two groups of students with regard to satisfaction, learning process, and achievement were analyzed.

Keywords

Online collaborative learning, Satisfaction, Academic performance, Knowledge construction, Cultural context

Introduction

A growing amount of research focuses on learning in group settings and more specifically on learning in computer-supported collaborative learning (CSCL) settings. Interactive technologies, such as web-based technology, can enhance the collaboration and construction of knowledge (Comeaux & McKenna-Byington, 2003). Group discussion, during which students develop effective cognitive learning strategies through social interactions, is one of the key activities of collaborative learning. These learning strategies encourage the adoption of a deep learning approach and have been shown to be effective in enhancing student achievements (Garrison & Cleveland-Innes, 2005).

In CSCL, learners are encouraged to exchange ideas, share perspectives, and use previous knowledge or experience in order to decide on the best solution for problems (Dewiyanti, Brand-Gruwel, Jochems, & Broers, 2007). Previous studies confirm that student involvement is more intense and equally distributed among group members in CSCL environments compared to face-to-face sessions (Angeli, Valanides, & Bonk, 2003). Recent studies indicate that online collaboration such as asynchronous discussions enhances student learning achievement (Young, 2008).

Culture serves as a perceptual framework that guides the interpretation of interactions and the construction of meanings (Berry, Poortinga, Segall, & Dasen, 2002; Cortazzi, 1990). Cultural attributes can affect online presence and learner perceptions. It is important to consider the cultural backgrounds of learners if we are to understand how they respond to computer-based learning (McLoughlin & Oliver, 2000). Some previous studies have indicated cultural gaps between “Confucian-heritage” and “western” learners in online collaborative learning environments, however, mostly in western educational settings. Few empirical studies have focused on comparing student attitudes, behaviours, and performance in Confucian-heritage Chinese educational settings and western educational settings. This study aims to study online collaborative learning in Flemish and Chinese contexts, with the former being more of an individualist culture, and the latter more of a collectivist culture. The study focused on the investigation of student satisfaction, performance, and knowledge construction through online collaboration in the two different cultural settings.

Student satisfaction with online collaborative learning and preferences across cultural contexts

When an e-learning environment is applied, student satisfaction should be considered in evaluating the effectiveness of e-learning. The degree of student learning satisfaction with an e-learning environment plays an important role in the adoption of e-learning or blended learning. Learners’ satisfaction can have repercussions on whether learners like to use systems or not, how learners work together, and whether there is a good working atmosphere among learners.
Some studies found that students who participated in online collaborative tasks expressed higher levels of satisfaction with their learning process compared to students who didn’t participate in online collaborative learning (Jung, Choi, Lim, & Leem, 2002). Additionally, in web-based collaborative learning systems, learners’ satisfaction with collaborative learning can be described as the degree to which a learner feels a positive association with his/her own collaborative learning experiences (Dewiyanti et al., 2007). Studies have indicated that students from one cultural context may have different attitudes towards educational interventions that are based on practices in another cultural context (Chang & Tsai, 2005). Therefore, more comparative research is needed regarding learners’ interaction online and the impacts of cultural differences on student online collaboration (Kim & Bonk, 2002). In addition, student satisfaction in e-learning environments is a critical issue and has been questioned in some research (Santhanam, Sasidharan, & Webster, 2008; So & Brush, 2008; Wu, Tennyson, & Hsia, 2010).

This study involves students in two different cultural contexts. The Flemish culture is situated in a western setting, which is more individualistic, while the Chinese culture, as part of the Confucian-heritage cultures, is traditionally representative of a collectivistic culture (Baron, 1998; Hofstede, 1986). Previous studies indicate that students from Confucian-heritage cultures (CHC) show a tendency to be introverted and passive, and less active in online collaboration. Warden, Chen, and Caskey (2005) observed that students from CHC rely on teachers to guide study strategies. Research by Chen (2010) indicates that online interactions among CHC students are largely confined to an instructivist approach on the part of the teacher. A study by Smith, Coldwell, Smith, and Murphy (2005) found that Chinese students were significantly less comfortable with discussions in e-learning compared to western students. They also found that Chinese students contributed less to online discussions. These results show that there are distinct features in online collaborative learning experiences, participation, and satisfaction of students from different cultural backgrounds.

Computer-supported collaborative learning and knowledge construction

Collaborative learning is a social interaction that involving a community of learners and teachers, where members acquire and share experience or knowledge. Computer-supported collaborative learning (CSCL) is based on the pedagogical assertion that students learn and construct knowledge through group interaction (Puntambekar, 2006). Collaborative learning involves the joint construction of meaning through interaction with others (Law & Wong, 2003). CSCL promotes meta-cognitive processes, reflective interaction, and problem solving (Jonassen & Kwon, 2001). In a constructivist learning environment, students are more interested, and critical thinking and inquiry is promoted (Mayes, 2001). Educational research has shown that more effective learning takes place if learners are actively involved, rather than being passive listeners (Nurmela, Palonen, Lehtinen, & Hakkarainen, 2003). Based on social constructivism, working together while accomplishing a task is seen as a characteristic of a powerful learning environment, which facilitates active construction of knowledge (Van Merrienboer & Paas, 2003). Studies found that students in collaborative learning conditions had more constructive learning processes (Eichler, 2003). CSCL can lead to the successful development of learning improvement and learners’ knowledge sharing and knowledge construction (Walker, 2005). In online learning communities, students can create, share information, practice critical reflection, negotiate meaning, test synthesis, and build consensus. Through online, collaborative written assignments, group discussions, debates and critiques of arguments, students can enhance knowledge construction.

With regard to content analysis of online interaction and discussions, the analysis model of Veerman and Veldhuis-Diermanse (2001) builds on social-constructivist principles. It focuses on two main discussion behaviours, namely task-oriented and non-task-oriented communication. The model of Gunawardena et al. (1997) proposes a model for evaluating the construction of knowledge through social negotiation. The five phases include sharing and comparing information, exploring dissonance, negotiating meaning, testing synthesis, reaching and stating agreement, and applying co-constructed knowledge.

Student performance in online discussions and group work

Student performance in online learning is emerging as a crucial ingredient in the evaluation of online learning environments. Student learning experience, the learning context, and the learning outcomes are not to be seen as separate variables and processes (Prosser & Trigwell, 1999). Empirical studies reveal a positive correlation between
students’ visible learning behaviours, such as participating in online activities, and their learning outcomes (e.g., Wang, 2004). However, there were very limited empirical studies examining student learning processes and outcomes in distinct cultural and educational settings. Previous studies mostly compared student performance in online learning and traditional learning (e.g., Stansfield, McLellan, & Connolly, 2004). In this study, we compare the performance in online discussions and group work between the Chinese and Flemish student groups.

Research questions

The following research questions are put forth for this study: 1) What are the differences in the satisfaction with the online collaborative learning between the Chinese and Flemish students? 2) What are the differences in student learning performance between Chinese and Flemish groups? 3) Are there cultural differences in the level of student knowledge construction through social interaction in online discussions?

Method

Research setting

The current study focused on examining the satisfaction, online performance, and knowledge construction through peer interaction of students in different cultural contexts. For this purpose, a parallel e-learning platform and course design was set up in both a Flemish university and a Chinese university. The e-learning platform is an open-source platform based on Dokeos. Efforts were made to make the learning design as similar as possible in the two educational settings. The same lectures were presented and the same online tasks were assigned to both the Chinese and Flemish groups during one academic semester. The student groups were in a similar age range and with similar subject knowledge background, as both were first-year university students. The online tasks were designed in a way that was suitable for both target groups. The online discussion and group work centered on themes in educational sciences. In view of each theme, authentic tasks were presented to the groups of students. For example, for the theme “constructivism,” students were required to create wiki pages to explain the main concepts in a brief way and give examples. Students were able to use different sources such as articles, books, websites, photos, newspapers, and audio/video fragments to explain the different elements theoretically as well as to provide examples. They also needed to try to make the wiki attractive/inviting for readers. Students were divided into groups of six members. Students were trained on how to use the e-learning system, how to participate in group discussions, and how to create wiki documents and pages.

Participants

For this study, 163 Chinese students and 208 Flemish students majoring in educational sciences were involved. The Chinese students \( (n = 163) \) were from a major comprehensive university in Beijing. The average age of the Chinese students was 19.3. The Flemish students \( (n = 208) \) were from a comprehensive Flemish university in Flanders. Their average age was 19.8.

Procedure and instruments

In both the Flemish and Chinese settings, students were randomly assigned to a group of six students. After each theme of lectures, students were required to participate in the online group discussions and group work on an assignment. Each online assignment lasted two weeks, and the students were required to contribute to online discussions and group work at least twice a week. Two teaching assistants were assigned as supervisors for each of the Flemish and Chinese student groups. After three months of online work, student online contributions were assessed on the basis of qualitative and quantitative criteria that were communicated to the students at the start of the course. Assessment was based on group achievement, and each group got a score for their online performance.
Questionnaire on student satisfaction and dissatisfaction with online collaborative learning

After one semester of online collaborative learning, students were asked to fill in a questionnaire about their satisfaction and dissatisfaction with online collaborative learning. This questionnaire consisted of 15 questions assessing the satisfaction of collaborative learning and 15 questions for students to choose and rank the aspects that they like or dislike most. Students were required to report on a Likert scale (from 0 to 6) the extent that a certain statement was true or false or the extent to which they liked or disliked a certain function of the e-learning environment. Student satisfaction reflects five dimensions: e-learning function, collaborative learning, peer contribution, interaction, and group results. Sample questions include: “I am happy that I can work together with others on the assignments”; “I am happy that working together with others helps me gain a deeper understanding of the course content”; “Working online with group work (such as wiki) is new and exciting for me”; and “I am satisfied that each member of my group equally contributes his/her part in the group assignments.” The psychometric quality of this measurement was confirmed with a Cronbach’s $\alpha$ score higher than .75. In addition, the students also reported their demographic features and the average time they spent on the online discussion and group work per week.

Content analysis

The scripts of each group of Flemish and Chinese students were coded and analyzed. The data set comprises the transcripts of all messages posted during group discussions by these groups during one semester. We applied the coding scheme of Veerman and Veldhuis-Diermanse (2001) to analyze the distribution of communication types and the coding scheme of Gunawardena, Lowe, and Anderson (1997) to analyze the level of social construction of knowledge. In our research, the complete message was used as the unit of analysis. When a message comprised elements of two different levels of knowledge construction, the highest level was assigned. The messages in the transcripts were coded by three independent coders for the Flemish students and three for the Chinese students. The Chinese and Flemish coders received training by the same researcher to get acquainted with the coding schemes using the same sample data. The inter-rater reliability was checked by determining percent agreement between the raters. For the raters of the Flemish group, the percent agreement was .91; for the raters of the Chinese group, the percent agreement was .86.

Statistical analysis

$T$-tests were used to analyze the differences between the Chinese and Flemish students regarding their satisfaction and dissatisfaction with the online collaborative learning. Chi-square analysis was adopted to compare the student message types and the level of knowledge construction. Furthermore, the achievements of Chinese and Flemish students in online group assignments were compared. With regard to content analysis of online group discussions, two coding schemes (Veerman & Veldhuis-diermanse, 2001; & Gunawardena et al., 1997) were used in this study.

Results

Student satisfaction and dissatisfaction with online collaborative learning

Significant differences were found between Chinese and Flemish students regarding their satisfaction and dissatisfaction with online collaborative learning. The Chinese students reported a higher level of satisfaction with the e-learning functions, online collaboration, and peer contribution compared to the Flemish students ($p < .05$). Compared to the Flemish group, the Chinese group was more satisfied with the equal contribution of group members ($p < .01$). In addition, the Chinese group preferred working together with others on the assignments than did the Flemish group ($p < .01$). Chinese students also reported to a larger extent that the online collaborative learning is “new and exciting” compared to the Flemish group. The Flemish students were more satisfied with the final results of the online group work compared to the Chinese group ($p < .001$), and they spent more time in average on the online group collaborative learning, 4.85 hours per week versus 2.26 hours per week for the Chinese students. With regard to the dissatisfaction of students, the Chinese group more often reported a lack of interaction between students and teacher in asynchronous group discussions compared to the Flemish group. The Flemish group reported to a
larger extent that working on the tasks online was time-consuming compared to the Chinese group \((p < .001)\). The Chinese students were less happy with task division in online group work than the Flemish students. The main differences between the two groups are summarized in Table 1.

**Table 1. Student satisfaction and dissatisfaction with online collaborative learning**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Sig.(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction with online collaborative learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with collaborative learning in the e-learning environment</td>
<td>3.65</td>
<td>3.24</td>
</tr>
<tr>
<td>Satisfaction with peer interaction within the group</td>
<td>3.60</td>
<td>3.56</td>
</tr>
<tr>
<td>Satisfied with the functions of the e-learning environment</td>
<td>3.76</td>
<td>3.37</td>
</tr>
<tr>
<td>Satisfied with the final results on the group assignment</td>
<td>3.85</td>
<td>4.66</td>
</tr>
<tr>
<td>Satisfaction with equal group member contribution</td>
<td>3.43</td>
<td>2.83</td>
</tr>
<tr>
<td>Satisfaction with the opportunity that group members can work together on assignments</td>
<td>4.29</td>
<td>3.83</td>
</tr>
<tr>
<td>Satisfaction that working together can help me gain a deeper understanding of the course content.</td>
<td>4.48</td>
<td>3.64</td>
</tr>
<tr>
<td><strong>Dissatisfaction with online collaborative learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-consuming</td>
<td>3.24</td>
<td>4.50</td>
</tr>
<tr>
<td>Dissatisfaction with task division</td>
<td>2.76</td>
<td>2.25</td>
</tr>
<tr>
<td>Lack of interaction with teacher</td>
<td>4.09</td>
<td>3.43</td>
</tr>
</tbody>
</table>

\* \(^p < .05\), ** \(^p < .01\), *** \(^p < .001\).

\(^a\) Adjustment for multiple comparisons: Bonferroni correction is applied

Despite the differences, similarities between Flemish and Chinese students were also found. Both Chinese and Flemish students reported that it was an advantage to be able to work at their own pace, and found it was an advantage that each group member could contribute to the group assignments in online collaborative learning. Both Chinese and Flemish students reported that online collaborative learning helped them to gain more knowledge than if they would have studied alone. They also stated that they had learned a lot, considering the time they’ve put into the online collaborative learning assignments. The Chinese and Flemish students were similarly satisfied with the peer interaction and with the technical help they received from the course coordinators.

As to what the students were satisfied and dissatisfied with, we found that the Flemish students liked working at their own pace most of all, while the Chinese students best liked the fact that they could work together with others on the assignments. What the Flemish students disliked most was that working on the tasks online was time-consuming, whereas the biggest problem the Chinese group reported was the lack of interaction between students and teacher.

**Student online learning performance and group achievement**

Students’ online group work was assessed by two teachers in each context. One teacher, who is bilingual, assessed both student groups; and a second teacher in the Flemish and Chinese context assessed the work of Flemish and Chinese student groups separately. The assessment criteria were the same for both settings and were communicated beforehand to all teachers and students involved. The assessment was based on the frequency of group-member contributions to the assignment (the system can track all member contributions for each task), and the quality of the group work. For each task, each group got a score; and the final group score was the average score for all group assignments. Student group achievements were compared, and the results show that, in general, the Flemish students had a slightly higher mean score than the Chinese students (Table 2).

**Table 2. Selected student group scores of online group work**

<table>
<thead>
<tr>
<th>Group achievement</th>
<th>Mean (SD) of assignment score (out of 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Flemish groups</td>
</tr>
<tr>
<td>G1</td>
<td>59 (11)</td>
</tr>
<tr>
<td>G2</td>
<td>57 (11.6)</td>
</tr>
<tr>
<td>G3</td>
<td>50.5 (13)</td>
</tr>
</tbody>
</table>
Content analysis of student knowledge construction through social interaction

In average, the Flemish students posted weekly more messages per person (7.5 messages) in asynchronous group discussions compared to the Chinese students (3.9 messages). For both groups, there were no significant differences as to the number of messages posted by male and female students. To test whether the types of messages and the achieved level of knowledge construction differ significantly, chi-square analyses were applied. The distribution of types of message and level of knowledge construction through social negotiation of the two groups are presented in Table 3.

The types of messages posted by both groups were rather similar, with a majority of them being task-oriented messages. The two groups of students seemed to be similar regarding non-task-oriented messages, which were technical, social, or related to planning. With regard to the levels of knowledge construction, the pattern of both groups was also similar. Both groups contributed a majority of messages that were at the first level of knowledge construction: sharing and comparing information. However, Flemish students contributed a higher frequency of messages that were at the second level of knowledge construction, exploration of dissonance, compared to Chinese students. Both groups contributed to a similarly lesser extent (about 12%) messages that were at the third level of knowledge construction: negotiation of meaning. Both groups contributed very few messages (less than 4%) that reached the fourth and fifth levels of knowledge construction.

<table>
<thead>
<tr>
<th>Types of messagesa</th>
<th>Chinese</th>
<th>Flemish</th>
<th>$x^2$</th>
<th>$p^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task oriented</td>
<td>95.9%</td>
<td>94.2%</td>
<td>0.35</td>
<td>.505</td>
</tr>
<tr>
<td>Non-task oriented</td>
<td>4.1%</td>
<td>5.8%</td>
<td>11.25</td>
<td>.006</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>0.3%</td>
<td>1%</td>
<td>11.58</td>
<td>.005</td>
</tr>
<tr>
<td>Technical</td>
<td>0.1%</td>
<td>0.5%</td>
<td>10.16</td>
<td>.052</td>
</tr>
<tr>
<td>Planning</td>
<td>0.7%</td>
<td>1.2%</td>
<td>10.11</td>
<td>.053</td>
</tr>
<tr>
<td>Social</td>
<td>3%</td>
<td>3.1%</td>
<td>0.27</td>
<td>.641</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Levels of knowledge constructionb</th>
<th>Chinese</th>
<th>Flemish</th>
<th>$x^2$</th>
<th>$p^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sharing and comparing information</td>
<td>79.5%</td>
<td>63.7%</td>
<td>1.65</td>
<td>.121</td>
</tr>
<tr>
<td>2. Exploration of dissonance</td>
<td>5.4%</td>
<td>19.7%</td>
<td>50.32</td>
<td>.000</td>
</tr>
<tr>
<td>3. Negotiation of meaning</td>
<td>11.3%</td>
<td>12.6%</td>
<td>3.88</td>
<td>.054</td>
</tr>
<tr>
<td>4. Testing synthesis</td>
<td>1.7%</td>
<td>2.8%</td>
<td>15.77</td>
<td>.005</td>
</tr>
<tr>
<td>5. Agreement statements and applications of newly constructed meaning</td>
<td>2.1%</td>
<td>1.2%</td>
<td>12.10</td>
<td>.006</td>
</tr>
</tbody>
</table>

b. Coding based on Gunawardena et al., 1997.
c. Adjustment for multiple comparisons: Bonferroni correction is applied.

Discussion

This study focused on three key issues in relation to student satisfaction with the online learning environment, their online performance, and knowledge construction in online group discussions.

Surveying students’ satisfaction with collaborative e-learning is a critical issue in promoting the innovative use of modern educational technology, especially in different cultural contexts. Our results indicate that there were significant differences between Chinese and Flemish students regarding their satisfaction and dissatisfaction with
online collaborative learning. On average, the Flemish students spent more time in online collaboration and were more satisfied with the results of group work than were the Chinese students. The Chinese students enjoyed online collaborative learning to a greater extent and were happier with the contributions of group members than were the Flemish students. Both groups of students were satisfied with the functions of the e-learning environment, appreciated the opportunities to work collaboratively, and agreed that collaborative learning promotes deeper understanding of the learning content. The results are consistent with previous studies that students in general are satisfied with online collaborative learning (Dewiyanti et al., 2007). The Flemish students ranked flexibility in time as the main advantage of e-learning, and the Chinese students found that working collaboratively online was a big advantage. Both groups of students were positive about working on a group product.

As to student dissatisfaction, the Chinese students found that the lack of teacher guidance and interaction in the e-learning environment was the biggest problem for them. Although teacher guidance was at about the same level for the Flemish students, the latter group found it less of a problem. This might be due to the different expectations of teacher involvement with the two distinct groups. Teachers or tutors play a very important role in Chinese educational contexts. Observations of the current e-learning programs in China indicate that e-learning tends to be heavily instructor-centered, for example, using video lectures online. Other studies comment that Chinese e-learners found it problematic when the teacher or tutor presence was low (Friesner & Hart, 2004). This could also be because of the low ambiguity tolerance of Chinese students who expect the presence of expert and certain knowledge (Zhu, Valcke, & Schellens, 2008a), which leads to a stronger need for feedback and teacher help in the learning environment (Anderson, 2000). The new and exciting online collaborative learning approach did not result in more intensive involvement of the Chinese participants; they were less active than Flemish students in terms of the time spent online and the messages posted. This might be because the Chinese students were less familiar with this type of learning approach than were the Flemish students. It might also be related to the fact that Chinese students did not have as easy access to computer and Internet and were less familiar with computer use compared to Flemish students, as identified in an earlier study in similar settings (Zhu, Valcke, & Schellens, 2008b). Flemish students rated “time-consuming” as the primary problem, but most likely, thanks to their extensive participation, they were quite satisfied with their final results in the group work. Another negative effect was the technology dimension, which was reported as the second problem by both groups of students. This is not surprising for new learners in e-learning, but attention should be paid in the future to provide more appropriate training and technical support to students (Fallshaw & McNaught, 2005). Our results were consistent with a previous study by Smith et al. (2005) that Chinese students were significantly less comfortable with discussions in e-learning compared to western students. They also found that Chinese students posted fewer messages to the online discussions. These results point out there are distinct features in the online collaborative learning experience, participation, and satisfaction of students from different cultural backgrounds.

With regard to student performance in online group assignments, the Flemish students performed better than the Chinese students on the group assignments. This might be related to the more intensive involvement of the Flemish students. In addition, Flemish students seemed to be more used to questioning and expressing different ideas. However, the Chinese students were more reluctant to respond directly to the views of others or express different opinions. This might have influenced their final group performance.

Students’ perceived satisfaction and their performance in online collaborative learning are important factors to determine whether an innovative learning approach can be applied in a sustainable way. Our study confirms that there are significant cultural differences in student satisfaction and academic achievement in an innovative e-learning environment.

Regarding student knowledge construction through online discussions, Chinese students posted relatively fewer non-task oriented messages than did the Flemish students, but for both groups, the majority of group communications were task-oriented. Activities such as asking, arguing, explaining, and providing extra resources dominated the discussions. These findings are in line with previous studies on online collaborative learning (Schellens & Valcke, 2006). At the knowledge-construction level, the results show that for both Chinese and Flemish students, a majority of messages have been coded as level 1 (sharing and comparing of information). However, Flemish students posted more messages of level 2 (exploration of dissonance). This might be due to the fact that Chinese students did not want to openly disagree with their fellow group members. In addition, because dissonances and disagreements were expressed more subtly by Chinese students, it was more difficult to classify the messages. Moreover, there were fewer messages reaching the higher levels of knowledge construction. This distribution of students’ contributions
across the five levels of knowledge construction corresponds with previous findings that few messages reach the fourth and fifth level and that a majority of messages are at the first level (Gunawardena et al., 1997; McLoughlin & Luca, 2002). This could be explained by the learning habits of students. Students, especially freshmen, were not yet used to testing syntheses, summarizing agreements, and applying newly constructed knowledge. They applied more often the first level of knowledge construction, which is a prerequisite for a discussion and for maintaining the flow of interaction. These contributions are in a way indispensable in order to elicit contributions at a higher level of knowledge construction. Because the discussion tasks were new to students for each theme in our study, we did not expect significant differences between the discussion themes. Related studies of De Wever, Van Keer, Schellens, and Valeke (2007) involving content analyses of students’ asynchronous discussions in similar Flemish setting indicate that there was no gradual increase of students’ level of knowledge construction throughout the different discussion themes because the discussion tasks of each theme were new to students.

**Limitation and conclusions**

It has to be noted that the results should be considered in a cautious way as the study is applied in specific settings. The findings of this research may only be applicable in similar contexts. It also has to be pointed out that although we have identified a series of differences and similarities between the two cultural groups, individual differences should not be neglected. Furthermore, the differences in the results of the two settings can be explained not only in relation to cultural differences, but also in relation to the new instructional experience for the Chinese students. In addition, although we tried to control several educational setting variables, we realize that other variables might exist, such as social and economic environment, educational systems, and campus environment, which might have influenced student satisfaction, participation, and performance in the online collaborative learning environment.

At the content analysis level, quantitative content analysis was opted because of the large amount of messages. Future research could include more detailed and qualitative discourse analysis. In addition, the levels of knowledge construction might be influenced by the types of discussion tasks and structuring support, which could be examined in following studies. Other coding schemes could be used in future studies.

In conclusion, this study confirms that there are significant cultural differences in student satisfaction, academic performance, and knowledge construction in an online collaborative learning environment. It also indicates that students’ perceived satisfaction and their performance in online collaborative learning are important factors to determine whether an e-learning approach can be applied in a sustainable way. Furthermore, the study indicates that learning with peers may benefit not only the overall individual performance, it may also enhance team performance by increasing the quality of team product. Students can learn to formulate ideas and opinions more effectively through group discussion. Based on social constructivism and activity theory, the online learning system can enrich collaborative learning activities for knowledge construction. Results of this study confirm that online learning system can enrich students’ collaborative learning activities and their knowledge construction via group interaction.

Finally we gain insights from this study that culture is an important variable to consider with regard to instructional design in different cultural contexts. Student satisfaction with and the level of knowledge construction in the e-learning environment are also important variables influencing student learning, especially in a student-centered e-learning environment. Understanding these variables would be helpful for instructors to design meaningful educational activities to enhance student satisfaction and performance and to promote student knowledge construction through social and peer interaction.

**References**


Design of a Motivational Scaffold for the Malaysian e-Learning Environment

Nor Aziah Alias
Faculty of Education, Universiti Teknologi MARA, 40200 Shah Alam, Selangor, Malaysia // noraz112@salam.uitm.edu.my

ABSTRACT
With a structure that provides control and choice over time, place, and pace, e-learning has emerged as a viable mode for working adults who wish to upgrade their knowledge. However, such flexibility provides leeway for these adults to procrastinate and for their motivation to dwindle. After all, many adults, especially Malaysian adults, are used to studying in a structured learning environment. It is thus necessary to scaffold these learners’ needs, particularly during the early stages of their learning endeavour. The purpose of this paper is to highlight the problem of sustaining the e-learners’ motivation in a Malaysian higher education setting. It also illustrates an approach taken to scaffold the learners’ regulation of motivation. Four-phase design and development research carried out to design, develop, and evaluate a web-based task support tool called the learning console is described. The paper also draws attention to the design principles upon which the affective component is built into the tool. Findings from the formative evaluation of the tool conducted among 40 participants consisting of experts, instructors, and e-learners suggested the capability of the tool to regulate the motivation of the learners. In addition, characteristics of both personalized and group learning were also apparent. The Learning console was found to scaffold the e-learners’ motivation and could be integrated into the existing learning management system.

Keywords
Regulation of motivation, Motivational scaffolding, Task support, e-Learning

Introduction
This paper focuses on e-learning in higher education. Most adults pursue e-learning courses as part of their lifelong learning effort and for the purpose of upgrading and updating their knowledge and skills. In this case, e-learning, is defined as “the use of Internet technologies to deliver a broad array of solutions that enhance knowledge and performance” (Rosenberg, 2001) and, hence, is used interchangeably with online learning. The challenge for online education providers not only centres on the issue of student recruitment, but also on how to retain them in the system once they have begun (Ludwig-Hardman & Dunlap, 2003; Jain, Abu, Akhbar, & Amat-Sapuan, 2004). Despite being adults, those trudging into the online learning environment still need support and structure to assist them in building their self-learning skills and to maintain interest and persistence. It is also necessary to help students assume responsibility and take initiative for their learning. This is especially significant among learners who have for most of their education been spoon-fed during regularly scheduled classroom sessions (Hisham, Dzakiria, & Walker, 2003; Teo & Gay, 2006; Andersson, 2008). In addition, isolation and alienation may lead to online learners’ feeling disconnected (Hara & Kling, 1999), to experiencing “psychological separation,” to be easily frustrated and distracted when studying on their own (Bauman, 1997), and to be less aware of what is expected of them, which inadvertently results in a last-minute preparation for exams and dissatisfaction. Procrastination is also a concern in a flexible online learning environment (Song & Hill, 2007; Tuckman, 2007). It is thus necessary to provide motivational support to learners in the online learning environment, especially in the early stages of immersion into an online learning environment. Support for students must not result in actions that reinforce their dependence. Hence, using and maximizing the benefits of technology, especially in the provision of scaffolds, was seen as more suitable for handling the problem (Boyer & Maher, 2004; Boyer, Phillips, Wallis, Vouk, & Lester 2008).

The main objective of this paper is to describe the design, development, and formative evaluation of the Learning console, a web-based task support tool meant to scaffold the motivation of Malaysian e-learners. The following questions were answered based on the specific objectives of the research:
1. Design objective: What strategies should be employed to regulate the motivation of the adult e-learners?
2. Development objective: How can web technology be used to scaffold the e-learner’s motivational self-regulation?
3. Evaluation objective: What is the adult e-learner’s reaction to the effectiveness, practicality, and value of a Learning console?
This paper reports only the early evaluation of the tool. It is not within the scope of this paper to provide the details of the summative evaluation.

**Scaffolding**

The concept of scaffolding is grounded in Vygotsky’s (1978) concept of assisted learning and learning potential as described by the zone of proximal development. The term *scaffolding* however, was introduced by Wood, Bruner, and Ross (1976) to mean tutoring or other assistance provided in a learning setting to assist students with attaining levels of understanding impossible for them to achieve without assistance. The idea of scaffolding is similar to having structures erected alongside buildings to support construction workers and later removed when the building is completed. Scaffolding involves providing learners with more structure during the early stages of a new learning venture and gradually turning responsibility over to them as they internalize and master the skills needed to engage in higher cognitive functioning (Palincsar, 1986 as cited in Ludwig-Hartman & Dunlap, 2003). Scaffolding is necessary to motivate learners and to aid them through the process of task attainment (Oliver & Herrington, 2001; Norton, 2005). Furthermore, scaffolded learning should eventually result in self-regulated learning (Winnips & McLoughlin, 2001). Scaffolding as a form of incremental support for learners aids them in organizing and monitoring their learning, as they complete different stages in the inquiry process. The learners can then be empowered to take control of their own learning.

The amount and type of scaffolding depends on the learners, the task, and the learning environment. It is the contention of the research that an online learner would benefit from scaffolding before he or she could take on the online learning environment on his or her own. Scaffolding may employ technology as in the case of an electronic performance support system (EPSS), or it can be strategies such as diagnostic pre-assessment, one-to-one advising or orientation to an online program (Ludwig-Hartman & Dunlap, 2003). As suggested by Oliver and Herrington (2001), scaffolding is essentially a teaching strategy that involves social interaction, discussion, and collaboration. The web has been identified as an excellent scaffolding tool due to its ability to provide devices such as email, threaded computer conferencing, chat rooms, hyperlinked resources, and collaborative workspaces (McLoughlin, Winnips, & Oliver, 2000). Winnips (2001) for instance, successfully constructed and evaluated the scaffolding-by-design model that was aimed at supporting student learning via web-based learning environments. The model was based on four aspects (task support, regulation, granularity, and management), which he termed as decision aspects. Three categories of scaffolds are inherent in the literature: (1) cognitive, (2) meta-cognitive, and (3) affective or motivational scaffolds. Cognitive and meta-cognitive scaffolds provide assistance, support, hints, prompts, and suggestions regarding the content, resources, and strategies relevant to problem solving and learning management, while motivational scaffolds involve techniques designed to maintain or improve the learner’s motivational state, such as attribution or encouragement (Azevedo, Cromley, Thomas, Seibert, & Tron, 2003).

**Motivational scaffolding in the design of the Learning console**

The majority of research in the area of scaffolding involves cognitive and meta-cognitive scaffolding (El Saadawi et al., 2010; Azevedo & Hadwin, 2005; McMahon, 2002; McLoughlin & Marshall, 2000; Cho & Jonassen, 2002; Cicognani, 2000). Research on motivational or affective scaffolding in the online learning environment is relatively scarce. There exist, however, notable efforts in the provision of this type of scaffolding in the work of Kim et al. (2006); Tuckman (2007); Boyer and Maher (2004); and Boyer, Phillips, Wallis, Vouk, and Lester (2008). Kim, Hamilton, Zheng, and Baylor. (2006), for instance, worked on a virtual peer while Boyer and Maher (2004) designed scaffolds to facilitate online adult learners’ social, self-directed learning. Tuckman (2007), on the other hand, studied the effect of scaffolding on the effectiveness of distance learning. Using chat rooms to run study skills support and instructor office hours, he found procrastinating students to perform better in a motivationally scaffolded environment. Tuckman’s scaffolding strategies are aimed at increasing the quality and frequency of learner-instructor and learner-learner interaction, with attention to feedback mechanisms, motivational meetings, and personal accountability. Balancing cognitive and motivational scaffolding by choosing a corrective tutorial strategy was found by Boyer, Phillips et al. (2008) to influence both student learning outcomes and self-efficacy gains.

It is imperative at this juncture to point out that this paper reports research that took an eclectic approach by employing cognitive-motivational scaffolds. Some aspects of Tuckman’s research were emulated in the design of a
scaffolding tool called “My Learning console,” which was aimed at enhancing the learner’s motivational self-regulation. It not only garnered the instructor’s role of facilitating and coaching but also employed technology to provide motivational elements in the form of intrinsic goals and extrinsic rewards. The design goals focused on the elements that mediate the interaction of an e-learner with the instructor in order to formulate his or her learning goals and choice of task. Apart from that, it also focused on how motivation is regulated through the utilization of the Learning console. The research had its theoretical underpinnings in the work of AlGhazali, an eleventh-century Islamic scholar; Boekaerts (1997); and Wolters (2003). Table 1 illustrates AlGhazali’s six steps. His concepts of **muhāsabah** (to note one’s strengths and weaknesses), **mujāhadah** (to strive and struggle to get things done), **murāqabah** (to perform one’s tasks in ways that is akin to having a higher authority watching), **mushāratah** (to set goals to achieve), **muṭāqabah** (to set consequences), and **muṭātabah** (to rebuke, to attempt to turn around) are translated into pragmatic guidelines that may sustain motivated learning behaviour.

The framework for the design of the scaffolding tool was further based on strategies by Boekaerts (1997), which were (1) goal setting, (2) maintaining action plan, (3) self-monitoring, and (4) self-evaluation. Strategies such as (1) self-consequating (students’ establishing and providing themselves with extrinsic consequences for different aspects of their engagement), (2) environmental control (students’ effort to arrange their surroundings to make completing a task easier), (3) interest enhancement (students work to increase their effort or time on task by making an activity more enjoyable), and (4) mastery and performance self-talk (emphasizing or articulating a particular goal or reason for wanting to complete the task), utilized by Wolters (2003) in his research, were also adapted. Table 2 shows the scaffolding strategies built into the design of the learning console, which are aimed at transforming environmentally regulated motivation to self-regulated motivation.

**Table 1. Al Ghazālī’s six steps (Adapted from UzmaMazhar, 2002, Six steps towards change)**

<table>
<thead>
<tr>
<th>Step</th>
<th>Arabic Term</th>
<th>English Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mushāratah</td>
<td>to make an agreement or contract (shart = stipulation). In this step, one must identify and set standards, conditions, limits, terms, and guidelines for the thoughts, feelings, and actions one is trying to achieve.</td>
</tr>
<tr>
<td>2</td>
<td>Murāqabah</td>
<td>to guard (raqab = guard). In this step, one must meditate before one’s actions. This requires that one think, contemplate, be introspective, and keep watch over one’s own self. We function as our own observer.</td>
</tr>
<tr>
<td>3</td>
<td>Muhāsabah</td>
<td>to evaluate self, taking account (hisab = account). This step involves self-examination in which one takes account of one’s own actions and continuously checks if one is upholding the agreement.</td>
</tr>
<tr>
<td>4</td>
<td>Muṭāqabah</td>
<td>to punish, to control (āqabah = punish). For the contract to work, we set consequences for ourselves when we have done something wrong and fail to keep the stipulations we agreed to uphold.</td>
</tr>
<tr>
<td>5</td>
<td>Mujāhadah</td>
<td>to try, strive (jahd = effort). In this step, one is fighting against one’s own lower self and inclinations. This is the stage of continuous and consistent struggle to overcome one’s desire.</td>
</tr>
<tr>
<td>6</td>
<td>Muṭātabah</td>
<td>to rebuke (atab = repent). In this step, if one has failed to maintain the contract, he makes the effort to turn around and change his ways upon recognizing the error.</td>
</tr>
</tbody>
</table>

**Table 2. Strategies to regulate the motivation of online learners**

<table>
<thead>
<tr>
<th>Scaffold: Environmentally regulated motivation → Self-regulated motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Elicit confidence through success opportunities.</td>
</tr>
<tr>
<td>▪ Reward competence (provide satisfaction).</td>
</tr>
<tr>
<td>▪ Make achievement and improvement explicit to the learner.</td>
</tr>
<tr>
<td>▪ Provide musyāratah (contract) murāqabah (guard), and muṭāqabah (consequence).</td>
</tr>
<tr>
<td>▪ Allow delving into relevant experiences to satisfy intrinsic needs or goals.</td>
</tr>
<tr>
<td>▪ Maintain diversity in task choice.</td>
</tr>
<tr>
<td>▪ Create an environment that enables reflection, self-consequating, self-monitoring, and self-talk.</td>
</tr>
<tr>
<td>▪ Encourage effort and persistence through continuous planning for task.</td>
</tr>
<tr>
<td>▪ Provide muḥāsabah (evaluate), mujāhadah (effort), muṭātabah (rebuke) and muṭāqabah (consequence).</td>
</tr>
</tbody>
</table>
Figure 1 illustrates how the abovementioned strategies fit into a generic self-regulation model (Grant, 2001) to support regulation of motivation of an adult online learner. The items in parentheses are elements that were not integrated in the scheme of this research. Nevertheless, their importance was not to be dismissed and should be considered for further research.

**Figure 1:** Motivation related elements and motivational regulation strategies built into the scaffolding tool

**Methodology**

The research methodology employed was grounded in the theoretical framework of design and development research (DDR) (Richey & Klein, 2007) that is theory driven, action oriented, interventionist orientated, participant centred, and collaborative. The features are similarly explained in the methodology of design experiment (Brown, 1992), development research (Van Den Akker, 1999), formative research (Reigeluth & Frick, 1999), design-based research (Bannan-Ritland, 2003), and educational design research (Plomp & Nieveen, 2007). A quick browse through recent publications in technology research shows an emerging trend in the adoption of this methodology in ICT research (Xie & Sharma, 2010; Shelton & Scoresby, 2010).

In DDR, the main research goal is to develop instantiations or approaches for solving human teaching and learning while at the same time, construct a body of design principles that can guide future development efforts (Reeves, 2000). In order to address the research questions, a functional prototype of the Learning console was developed and user-evaluated. The elements that characterized the DDR were as follows:

1. the intensive and systematic preliminary investigation of problems in the context of a university’s online distance learning environment
2. the systematic efforts in articulating the theoretical rationale for design choices
3. the delivery of clear empirical evidence about the validity, practicality, and effectiveness of the learning console for the target group in their settings and
Since the research sought to account for how the Learning console would function in real settings, it involved collaboration among the researcher, learners, and practitioners. It also required long-term engagement rather than a one-time treatment and survey. The major knowledge gained from the development research was in the form of design principles to support designers in their task (Van Den Akker, 1999). To summarize, the research was basically type I product design and development research aimed at the design, development, and evaluation of a specific product, namely, the learning console (Richey & Klein, 2007). However, the extensive review of literature carried out to establish the elements included in the design was distinctive of formative research (Reigeluth & Frick, 1999). Such effort contributed to the generalizability of the design and was seen by the researcher as providing directions to others who are confronting similar cases. On the other hand, this research was limited to what Bannan-Ritland termed as local evaluation rather than taking a more conclusive approach by including a broader impact evaluation.

There were four main phases in the research. The analysis phase was also an exploratory phase where problem diagnosis and articulation of theoretical foundation of the research took place. The researcher analyzed the learning environment, built a learner profile, re-examined her own experience, and studied relevant literature studied in order for solutions to be hypothesized. In the second phase, strategies were formulated, and the learning console was designed based on earlier analysis and prescribed guidelines. The third phase comprised the development of the tool based on earlier analysis, strategies, and prescribed guidelines. This phase involved extensive collaboration with the developer, users, instructors, and experts. The information derived from the analysis of data was used to enhance the appeal, value, effectiveness, and practicality of the learning console. The fourth phase consisted of the implementation and evaluation of the learning console by a group of online learners. During this phase, learners used the learning console in their natural setting, which was the online learning environment. Further changes were made based on the learners’ input, and a revised set of design principles was presented. Figure 2 illustrates the different phases through which the scaffolding tool is conceptualized and designed.

**Participants**

The participants in this study were learners, instructors, and experts in the field of online and distance education. The participants were sampled through a non-probabilistic mixed purposeful sampling technique (Patton, 2002), where the instructors and a subgroup of learners who met certain criteria were selected. The instructors and the reviewers were identified through their years of involvement in online distance learning. An associate professor who was a researcher in the field and serving as the instructional designer at the university was enlisted in a peer-review session to provide useful feedback on the practicality and for improving the interface design of the Learning console. She was selected due to her involvement right from the start of the university online learning program. Two Malaysian and two British experts who are highly respected figures in the fields of distance education, adult learning, online interaction, and motivation in online learning were identified via their extensive publication and sought-after expertise. They were contacted by phone and/or email; all of them agreed to review the validity and practicality of the learning console.

Several groups of online learners informed the research at the different phases. A group of ten was initially selected for the concept walkthrough, and another fifteen learners were selected for the small group evaluation. The selection of the instructor determined the selection of the group of learner-participants. This was to ensure that instructor factors such as inexperience would not impede learner’s experience when using the learning console. In addition, eight learners studying at a different university were also selected for the tryout.

The learners were first-year distance learners who had just encountered formal learning in an Internet-based environment. The choice of participants was in tandem with the concept of scaffolding propagated in the study. They were largely from Confucian-heritage societies that are known for their acquisition approach in study method and authoritative instructor’s role. The fact that Malays and indigenous learners make up the bigger proportion of Malaysian distance learners calls for the research to look at the cultural significance in the design of the online distance learning environment. Ziguras’ (2001) examination of the use of educational technologies to increase the flexibility of learning in Singapore, Vietnam, and Malaysia provided an account of the tendencies of Malaysian learners at three institutions to be spoon-fed, to expect more direction and closer supervision from locally available teaching staff, to “not want to be left alone,” and to have a high regard for lecturers. The authority and reverence of a
teacher/instructor continues to dominate the mindset of these learners (Churton, 2003), making them less independent. It is crucial to isolate the cultural traits that will enhance learning rather than dismissing all of them as deterrent to successful learning. AlGhazali’s six steps were adopted to address this particular issue.

**Figure 2:** The design and development research (DDR) processes

**Design and development of the learning console**

The first and second research questions were dealt with in the analysis, design, and development phase. As part of the design and development research agenda, a successful learning console was defined following an extensive literature review and survey of similar products. In addition, a survey was conducted on a group of distance learners, followed by several interview sessions. As a scaffolding tool, the learning console should meet the criteria of a scaffold. In this case, it meant a gradual transfer of responsibility and skills needed to complete particular tasks successfully (Brophy, 1999). It also meant that the learning console would allow the learners to self-manage the motivational aspects of task engagement (valuing, satisfaction) as they acquire the capacities for doing so.

The principles that guided the design of the learning console were then developed. Apart from being derived from theories of adult learning, learner-centred learning, motivation, and motivational self-regulation, the design principles also have their underpinnings in the researcher’s experience and common knowledge in the field of online distance learning. Four broad principles guided the design of the learning console: (1) A learner’s learning needs and existing knowledge must be linked to all future learning, (2) A learner’s cultural background and prior experience plays an eminent role in the learning process, (3) Motivational factors, such as intrinsic motivation, attributions for learning, and personal goals, as well as the motivational characteristics of the learning tasks, play a significant role in the learning process, and (4) The ability to monitor one’s own learning and regulate one’s motivation to learn is
necessary for success in an online distance learning environment. Further refinement and elaboration of the design principles are stated in Table 3.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Elaboration of principle</th>
</tr>
</thead>
</table>
| Embed guidance through negotiation, feedback and reminders | ▪ Encourage 24/7 interaction  
▪ Provide timely feedback and reminders  
▪ Utilize learner-instructor interaction as a prompt for learning |
| Foster motivational self-regulation            | ▪ Create an environment that enables reflection, self-consequating, self-monitoring and self-talk  
▪ Provide *mushāratah, muhāsabah, mujāhādah, murāqabah, mu'ātabah* and *mu'āqabah*  
▪ Encourage effort and persistence through continuous planning for task |
| Create safe, inclusive, and respectful environment | ▪ Provide flexibility; non-compulsory use of the support tool |
| Keep learning engaging and motivating          | ▪ Allow delving into relevant experiences to satisfy intrinsic needs or goals  
▪ Maintain diversity in task choice  
▪ Elicit confidence through success opportunities  
▪ Reward competence (provide satisfaction)  
▪ Make achievement and improvement explicit to the learner |
| Provide adequate scaffolding                   | ▪ Maintain balance between instructor control and total learner reliance  
▪ Play down reverence of the instructor by gradually transferring learning responsibility to the learner |
| Accommodate positive cultural elements and prior learning experience | ▪ Capitalize on multiple goals approach  
  ▪ Use extrinsic motivators and performance goals with a focus on high achievement  
  ▪ Tie reward to mastery goals  
▪ Emphasize assessments  
▪ Support individual learning effort  
▪ Support instructor’s role as both a coach and a guide  
▪ Acknowledge learner’s needs of instructor guidance and supervision |

**Description of the learning console**

The learning console is a web-based task support tool. It has three main features: (1) task negotiation, (2) point accumulation status, and (3) task reminder. They are incorporated in the learning console as “my negotiated task,” “my status,” and “my message box.”

![The three main components of the learning console](image)

The first feature provides the means for students to negotiate tasks with the instructor instead of working on instructor-assigned tasks. The second feature provides a means for students to track his/her current academic status in the course (at any time) in which a student’s current coursework is exhibited in a time-based graph. Points given to tasks negotiated are cumulatively and constantly added to a student’s grade, and the progress is depicted graphically.
The instructor plays a significant role in ensuring that the learner completes his or her task. Feedback and comments from the instructor are retrievable from the message box, and hyperlinks are created on the “my status” page. Individual reminders in the form of web messages, emails, or messages via short messaging service will also be sent to the learners should they fall behind in their tasks. Table 4 describes the features and the associated strategies formulated to scaffold the learner’s motivational self-regulation.

Table 4. Features of the learning console

<table>
<thead>
<tr>
<th>Feature</th>
<th>Scaffolding strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task negotiation</strong></td>
<td>• Provides space where the learner can click on options to start a task, proposes topic, provides description of the task, selects the resources to use and the format of final product, and proposes schedule</td>
</tr>
<tr>
<td></td>
<td>• Displays negotiated and approved task and those that are still pending</td>
</tr>
<tr>
<td></td>
<td>• Provides means for learner to keep negotiating task in order to improve academic status</td>
</tr>
<tr>
<td></td>
<td>• Displays assessment of completed task (marks, comments, recommendations)</td>
</tr>
<tr>
<td></td>
<td>• Provides space for feedback from instructor</td>
</tr>
<tr>
<td><strong>Point accumulation status</strong></td>
<td>• Displays a learner’s current accumulated points gained from completed tasks</td>
</tr>
<tr>
<td></td>
<td>• Provides the learner with a sense of achievement</td>
</tr>
<tr>
<td></td>
<td>• Provides means for a learner to view feedback or comments from the instructor on the completed task</td>
</tr>
<tr>
<td></td>
<td>• Provides impetus to improve or to raise accumulated points</td>
</tr>
<tr>
<td><strong>Task reminder</strong></td>
<td>• Provides instructor with the means to probe and remind learners via message box or emails</td>
</tr>
</tbody>
</table>

Figure 4. Flow chart of the learner interaction with the learning console

**Flowchart of the interaction**

To understand how a learner may use the learning console, a flow chart of the process is presented in Figure 4. Figures 5, 6, 7, 8, and 9 further illustrate the features of the learning console (http://mylearningconsole.com). Since the learning console was developed as part of a larger learning management system (LMS), it is assumed that other
aspects of online learning, such as peer collaboration, online learning resources, and cognitive and meta-cognitive support, were readily built into the learning environment.

Figure 5. Home page

Figure 6. Task negotiation panel on My Negotiated Task page

Figure 7: Task negotiated by learner
Data collection and data sources

As the intention was to gather information on users’ experiences in order to provide a basis for future use of the learning console as a support tool, in-depth semi-structured interviews with learners and instructors became the centre of the data collection process. Face-to-face interviews as well as email and phone interviews were conducted with the research participants. Additional data collected from a survey and session data such as session statistics and session patterns data were gathered to support the interview data and to enable the move from description to explanation. The survey which included thirteen five-point Likert scales anchored in strongly disagree and strongly agree items was developed corresponding to the design objectives that were validated during the expert reviews. Session statistics included the frequency and overall length in terms of time visited. Session patterns described the behaviour of students in a session. This pertained to the usage of the support elements, help sought, and messages sent during the session. The learning console interface was assessed by the experts based on a rating scale adapted from Reeves and Harmon (1993).

Several tryouts and user evaluation sessions took place before a group of fifteen learners were given access to the learning console. The learning console underwent revision and a second prototype was developed before the group of fifteen used it for the duration of a semester. The researcher employed an integrated approach that tested not just the product, but also all parts of the package that the users would receive, such as training, procedures, manuals, and online help. The basic questions that guided the revision of the learning console are illustrated in Table 5.
Table 5. Questions that guided revision of learning console

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Questions asked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity</td>
<td>Is the product an innovative element to support the Malaysian online learners?</td>
</tr>
<tr>
<td></td>
<td>Is the product consistent with the design objectives?</td>
</tr>
<tr>
<td>Practicality</td>
<td>Does the product run as intended?</td>
</tr>
<tr>
<td></td>
<td>Can the intended learners actually use the program?</td>
</tr>
<tr>
<td></td>
<td>Does the product appeal to the users?</td>
</tr>
<tr>
<td></td>
<td>Does the product provide opportunities for meaningful learner-instructor interaction?</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Does the product motivate learners?</td>
</tr>
<tr>
<td></td>
<td>Does the product assist the learners in directing their own learning?</td>
</tr>
<tr>
<td></td>
<td>Does the product support them in their learning process?</td>
</tr>
<tr>
<td>User friendliness, free from errors,</td>
<td>Is navigation clear to users?</td>
</tr>
<tr>
<td>efficiency, ease of use</td>
<td>Are the meanings of icons clear?</td>
</tr>
<tr>
<td></td>
<td>Do users get lost in navigating through the program?</td>
</tr>
<tr>
<td></td>
<td>Is the product efficiently constructed?</td>
</tr>
<tr>
<td></td>
<td>What is the minimal technology that it can be run on and do the target learners have access to that level of technology?</td>
</tr>
</tbody>
</table>

Formative evaluation of the learning console

The researcher conducted a qualitative evaluation during which the learning console was subjected to scrutiny by 40 research participants (experts, instructors, and learners) who were observed, consulted, and interviewed at four phases of the evaluation. The phases as promulgated by Flagg (1990) were (1) the planning and needs assessment phase; (2) the pre-production formative evaluation, which referred to the collection of information to guide decisions during the design phase; (3) the production formative evaluation, which referred to the gathering of data to guide revisions using the prototype; and (4) the implementation formative evaluation, which referred to testing the effectiveness of the product/program under normal use conditions with the intention of still making changes. The needs assessment phase was regarded as the first phase of evaluation by Flagg (1990), where pertinent information such as the learner profile was gathered.

Table 6 summarizes the different phases of the evaluation and the evaluation techniques associated with each phase.

<table>
<thead>
<tr>
<th>Phases of research</th>
<th>Phases of formative evaluation</th>
<th>Evaluation technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>Planning and needs assessment</td>
<td>Situational analysis (past research and document analysis)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Learner analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Learning environment analysis</td>
</tr>
<tr>
<td>Design</td>
<td>Pre-production evaluation</td>
<td>Interviews with learners and instructor</td>
</tr>
<tr>
<td>Development</td>
<td>Production evaluation</td>
<td>Concept walkthrough — analysis of group discussions</td>
</tr>
<tr>
<td>Implementation</td>
<td>Implementation evaluation</td>
<td>Peer review</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expert reviews</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One-to-one evaluation (learners and instructor)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In-depth interviews with learners and instructor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Survey</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online observation of students using the learning console</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expert consultation</td>
</tr>
</tbody>
</table>

Findings and discussion

The peer reviewer and the four experts generally agreed on the potential of the learning console to provide guidance and regulate motivation, to enhance self-evaluation, and to promote interaction and knowledge building among learners and the instructor. The instructors saw the learning console as an option to the existing mode of assessment
by turning over the responsibility to the learners and giving them the opportunity to improve throughout the semester of study.

The qualitative data provided a rich description of the users’ experience. The potential of the learning console to enhance learners’ motivational self-regulation was indicated by the learners who evaluated the learning console at all stages of the formative evaluation. They regarded the opportunity to continuously put in effort to improve and evaluate their learning process as a useful facet of the learning console. They were keen on the way they were able to negotiate their tasks and having them assessed and cumulatively added to their coursework. These features kept them motivated to continue learning. In addition, the learning console provided a sense of achievement as the learners were able to view their progress through the display of their accumulated points. These extrinsic benefits of the learning console that were constantly mentioned by the learners supported similar findings by Chang (2003) and Timmis and Cook (2001).

The survey indicated high mean scores for most of the questions. Learners rated highly the self-evaluation and regulation of motivation aspects of the learning console. They confirmed that the learning console had the potential to help them reflect on their learning process ($M = 4.0, SD = 0.47$), encourage their initiative to improve ($M = 4.33, SD = 0.67$) and self-evaluate ($M = 4.0, SD = 0.47$), keep them motivated to continue learning ($M = 3.9, SD = 0.74$), and provide a sense of achievement ($M = 3.8, SD = 0.79$). The learners also rated highly on the learning console’s potential to provide opportunities for enhanced learner-instructor interaction ($M = 3.9, SD = 0.74$). On the other hand, learners did not highly rate the learning console’s potential to provide feedback at key points ($M = 3.3, SD = 1.25$).

Learners who fully utilized the learning console apparently experienced intrinsic benefits that revolved around the freedom and opportunity to negotiate their ideas. Interestingly, the learners were mainly concerned about their own performance rather than that of others. Competition with other learners, as indicated by one learner, was not the motive for self-improving. Using terms employed by Kawachi (2003), the learner’s desire towards self-improvement can generate personal intrinsic motivation while the freedom to work on a task of his/her choice can lead to academic intrinsic motivation. Hence, the environmentally regulated motivation could lead to self-regulated motivation, should scaffolding be directed upon motivational traits such as challenge, interest, relevance, confidence, satisfaction, independence, and control.

The themes derived from the findings, however, illuminated a number of issues related to learners’ prior experience and characteristics. The instructor was the learner’s motivation to log in. Dependence on the instructor was still dominant, even though learners were adults entering their third semester of study. The groups of Malaysian learners were observed to need coaxing, showed a lack of exigency, unfortunately, and were generally pleased with the scores of their first test. Attempts to negotiate were observed later towards the end of the semester. Presupposing the technical and Internet competency of the learners was one of the assumptions of the research that fell short. Despite being in the online learning environment for some time, some of the learners were not comfortable with the web environment. What the researcher gained most from this research is the possibility of using the learning console to support both personalized learning and collaboration. This is undoubtedly a positive feature to deliberate on, in response to the different characteristics of the Malaysian online distance learners.

**Conclusions**

This paper illustrated research that included the design, development, and formative evaluation of a support tool to help solve the problem of sustaining motivation and enhancing the self-direction of online distance learners in a Malaysian context. The learning console and its features were considered by the research participants as having the potential to support both attributes; thus the design and development objectives were met. Although the level of use of the learning console was below expectation, the researcher discovered significant issues of prior practice and management of online learning that generated a revised set of design principles and implementation guidelines to inform others in similar settings that may be employed in future research.

Scaffolding should be addressed early in the stage of utilization since there is evidence for the need of different types and stages of scaffolding. The technical difficulties reported by some of the learners call for the provision of technology scaffolding and procedural scaffolding that consists of guides on how to use the learning console. The
lack of self-directing ability necessitates task scaffolding that can be implemented in stages, with the learners’ negotiating existing or instructor-assigned tasks and then proceeding to initiate their own tasks. Figure 10 exemplifies a simple framework for scaffolding the learner that can be further explored.

<table>
<thead>
<tr>
<th>Early stage</th>
<th>Intermediate</th>
<th>Later stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Induction to the learning console</td>
<td>Guided task negotiation</td>
<td>Self-initiated task negotiation</td>
</tr>
<tr>
<td>Types of scaffolding</td>
<td>Types of scaffolding</td>
<td>Types of scaffolding</td>
</tr>
<tr>
<td>• Technological</td>
<td>• Task-related</td>
<td>• Task-related</td>
</tr>
<tr>
<td>• Procedural</td>
<td>• Motivational</td>
<td>• Motivational</td>
</tr>
<tr>
<td>• Motivational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example of mechanism</td>
<td>Example of mechanism</td>
<td>Example of mechanism</td>
</tr>
<tr>
<td>• Walkthrough page</td>
<td>• Task reminder</td>
<td>• Task reminder</td>
</tr>
<tr>
<td>• Trial login</td>
<td>• Sample negotiated task</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 10. Suggested types and stages of scaffolding related to learning console use*

This research was the first iteration of a developmental cycle that encompassed a preliminary investigation and an active approach to design, redesign, and investigate constructs that are relevant to the researcher, learners, and instructors. Several iterations of data collection are recommended in order to allow the evaluation of the impact of the learning console and the learning outcomes of the learners. These include researching multiple sites and increasing the number of groups under different instructor management that will add to the dependability of results. Future work will deal with added functions of the learning console, such as the types of tasks negotiated, the effectiveness of group negotiations, the incorporation of other motivational elements, and the dynamics of learner-instructor interaction.

**References**


Oliver, R., & Herrington, J. (2001). Teaching and learning online: A beginner’s guide to e-learning and e-teaching in higher education. Edith Cowan University Publication, Western Australia.


Reeves, T. C. (2000, April), Enhancing the worth of instructional technology research through “design experiments” and other development research strategies. Paper presented at the International Perspectives on Instructional Technology Research for the 21st century, New Orleans.


Understanding of the Relationship Between Interest and Expectancy for Success in Engineering Design Activity in Grades 9–12

Oenardi Lawanto*, Harry B. Santoso and Yang Liu

Department of Instructional Technology and Learning Sciences, Utah State University, Logan, UT, 84322 USA // olawanto@usu.edu // harry.santoso@aggiemail.usu.edu // usuliuyang@gmail.com

*Corresponding author

ABSTRACT

In this study, the relationship between students’ interest in engineering design activities and their expectancy for success in grades 9–12 was evaluated. The theoretical frameworks developed by Eccles and colleagues (i.e., expectancy value) and Butler and Cartier (i.e., what the students bring to contexts under a self-regulated learning framework) were used to frame the relationship between students’ interest and expectancy for success. Specifically, this study extends the value part with interest by using psychological constructs of the Motivated Strategies for Learning Questionnaire (MSLQ). Three psychological constructs were used to evaluate students’ interest: intrinsic goal orientation (IGO), extrinsic goal orientation (EGO), and task value (TV). Two other psychological constructs were used to evaluate students’ expectancy for success in completing the design tasks: control of learning beliefs (CLB) and self-efficacy for learning and performance (SELP). A total of 113 students participated in the study. These students participated in five schools (in three states) that implement the Project Lead the Way curriculum. After finishing their design project, each student was asked to complete a modified version of the MSLQ survey instrument that evaluated the five aforementioned constructs. From a statistical test, it was found that there was a significant relationship between students’ interest in the engineering design tasks and their expectancy for success. Further statistical tests also indicated significant relationships between each construct of the interest and expectancy components. From a regression test, it was found that students’ intrinsic goal orientation and task value were significant predictors, each contributing about 55 and 34 percent, respectively, to students’ expectancy for success.

Keywords
Interest, Expectancy for success, Motivation, Engineering design

Introduction

It has been suggested by the Committee on K–12 Engineering Education (Katehi, Pearson, & Feder, 2009) that K–12 engineering education should emphasize engineering design. Everett, Imbrie, and Morgan (2000) noted that through the engineering design process “students not only know the mathematics and science but also actually understand why they need to know it” (p. 171). Thus, engineering design allows students to apply and integrate concepts and principles learned in other subjects in a more meaningful learning experience. Inasmuch as engineering design activities are promoted and predominant in most K–12 curricular and professional development programs, understanding the dynamics of students’ motivation in engineering design activities is critical. One common challenge facing engineering and technology educators is how to introduce and teach engineering design in a manner that would interest their students. This challenge involves cognitive and motivational aspects that impact students’ learning.

The human brain engages a wide range of cognitive processes to solve an ill-structured problem such as an engineering design task. It is well known that cognitive and motivational issues such as skill and will are interwoven (Corno & Mandinach, 1983). Studies have also suggested that motivation plays an essential factor in students’ learning accomplishments (Bong, 2001; Chemers, Hu, & Garcia, 2001; Chowdhury & Shahabuddin, 2007; Gore, 2006; Multon, Brown, & Lent, 1991; Zajacova, Lynch, & Espenshade, 2005). The objective of this study is to focus on the motivational aspects used by students in working on an engineering design task.

Numerous researchers have posited a variety of motivational constructs to explain how they affect achievement performance and choice. One is the expectancy-value model of achievement motivation (Eccles, Adler, & Meece, 1984; Eccles et al., 1983; Wigfield, 1994; Wigfield & Eccles, 1992, 2000). In it, the expectancies for success and ability beliefs, and task value are the major constructs studied by the researchers. Despite a number of studies that suggested the relationship between students’ motivation and learning performance, few studies have evaluated how the interest components of students’ motivation relate to their expectancy for learning success. Frick (1992) argued...
that interest influences “what people attend to, think about, discuss and learn more about” (p. 113). Understanding
the connection between how students’ interest in engaging in an engineering design task and their expectancy to
successfully complete the task should not only positively contribute to the knowledge building, but may also help
inform engineering- and technology-related educators in grades 9–12.

The term *expectancy for success* in this study is defined as one of the motivational components that leads an
individual to believe that he or she has the potential to successfully accomplish a task. This definition includes one’s
ability (i.e., self-efficacy), confidence (i.e., self-confidence), and control of his or her effort (i.e., control of learning
beliefs) to accomplish a task (Schunk, 1991; Shields, 1991; Wigfield & Eccles; 1992, Wigfield, 1994; Yong, 2010).
The *interest* in engaging in a task, on the other hand, is the motivational component that includes intrinsic and
extrinsic goal orientations that lead an individual to participate in a task and also his or her evaluation of how
interesting, important, and useful the task is (i.e., task value) (Geiger & Cooper, 1995).

This study was conducted to answer two research questions:
1) While students were engaged in an engineering design task, was there any significant correlation between the
   following?
   a) interest in the engineering design task and their expectancy for success
   b) intrinsic goal orientation and control of learning beliefs
   c) intrinsic goal orientation and self-efficacy for learning and performance
   d) extrinsic goal orientation and control of learning beliefs
   e) extrinsic goal orientation and self-efficacy for learning and performance
   f) task value and control of learning beliefs
   g) task value and self-efficacy for learning and performance
2) What is the relative importance of contribution of those interest components toward students’ expectancy for
   success?

**Relevant literature**

Numerous studies have revealed the relationship between expectancy value and academic achievement (Atkinson,
1957; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Eccles et al., 1983; Liew, McTigue, Barrois, & Hughes,
2008; Wigfield & Eccles, 1992; Wigfield, 1994). The relationship between students’ expectancy for success and
learning performance or achievement was also reported in a study conducted by Geiger & Cooper (1995). Similar
findings also confirmed that the students who had a high expectancy for success would achieve success, while those
who did not often failed in their tasks (Stephan, Bernstein, Stephan, & Davis, 1979). In these studies, students’
extpectancy for success was assessed by the students, and their academic achievement was evaluated by their
teachers. The way the students and teachers evaluated success may be different.

In one study, Cartier (2002) argued that what occurs in an individual’s learning is also influenced by what he or she
brings to each particular context. Strength, challenge, preference, and interest are important components of learning
(Butler & Cartier, 2005). It is important to investigate if students’ interest to engage in a task leads to their
expectancy for success. In this study, Pintrich, Smith, Garcia, and McKeachie’s (1991) motivational model was used.
Students’ interest was viewed from motivational constructs that include intrinsic goal orientation (IGO), extrinsic
goal orientation (EGO), and task value. Expectancy for success components includes self-efficacy for learning and
performance (SELP) and control of learning beliefs (CLB). Because interest and expectancy for success are both
viewed from the students’ perspectives, the results of this study offer potential for the improvement of engineering
and technology education in grades 9–12.

**Goal orientation**

Goal orientation in this case refers to the beliefs that induce one to approach, engage in, and respond to tasks in
different ways (Ames, 1992). There are two general categories of goal orientation: mastery, where students focus on
learning and mastering the material, and performance, where students demonstrate their abilities and performance
compared to others (Wolters, Yu, & Pintrich, 1996). In general, mastery goal orientation is associated with an
inclination for challenging tasks, an intrinsic interest in learning itself, and positive motivational beliefs such as high
self-efficacy. On the other hand, performance goal orientation is usually associated with seeking extrinsic rewards,
such as grades and evaluation by others, rather than an intrinsic interest in learning (Ames, 1992; Dweck & Leggett, 1988).

Performance goal orientation can be further divided into performance-approach, where students strive to achieve a goal and demonstrate their abilities, and performance-avoidance, where students’ major concern is avoiding failure and hiding their lack of abilities (Elliot, 1999). According to Yang, Tsai, Kim, Cho, and Laffey (2006), goal orientation can be subcategorized as intrinsic and extrinsic. Intrinsic goal orientation (IGO) involves the extent to which students perceive themselves as engaging in a task for reasons such as challenge, curiosity, and mastery. Extrinsic goal orientation (EGO) deals with the degree to which students perceive themselves as engaging in a task for reasons such as grades, rewards, performance, evaluation by others, and competition (Pintrich et al., 1991). Students with intrinsic goals are inclined to exert more effort and are more likely to persistent with difficult tasks than those with extrinsic goals (Ames & Archer, 1988; Merritte, 1999; Nicholls, 1984; Pintrich & Schunk, 2002). Although there are some differences in the literature about the categories of goal orientation, on the basis of the literature review and definitions above, we consider that mastery goal orientation is analogous to intrinsic goal orientation in that the two share some common constructs. To some degree, performance goal orientation is similar to extrinsic goal orientation.

**Task value**

Task value refers to a student’s perception of the extent to which the task is interesting, important, and useful. That is, what do students think of a certain task in terms of interest, importance, and utility? According to Wolters et al. (1996), mastery goal orientation is a positive predictor of task value and promotes adaptive motivational beliefs such as higher task value and self-efficacy. Also, task value is often positively related to self-efficacy. Both have been well documented as effective predictors of academic outcomes (Bong, 2004; Mutlon, Brown, & Lent, 1991). Students’ understanding about task is a key determinant of goal settings, strategy selection, and criteria used to self-assess outcomes (Butler & Cartier, 2005). The findings of Yang et al. (2006) suggested that students with high task value were more likely to improve their learning and performance by social interaction with peers.

**Self-efficacy for learning and performance**

Self-efficacy refers to judgments about one’s abilities to succeed in a given task (Bandura, 1997). It applies to a variety of contexts and is a good predictor of performance and behavior (Bandura, 1978, Gist & Mitchell, 1992). Numerous studies (e.g., Bandura & Schunk, 1981; Brown & Inouyne, 1978; Schunk, 1981; Weinberg, Gould, & Jackson, 1979) have suggested that strong self-efficacy is more likely to stimulate the exertion of greater effort to overcome a challenge, while weak self-efficacy tends to reduce one’s efforts or even cause a person to quit (Chowdhury & Shahabuddin, 2007). Self-efficacy influences how much effort is exerted, persistence when confronted with difficulties, and resilience to failures (Chowdhury & Shahabuddin, 2007).

**Control of learning beliefs**

Control of learning involves students’ beliefs that learning depends on their endeavors rather than external causes, such as the teacher. In this sense, if students believe that their efforts have a positive influence on their learning, they will be more likely to engage in learning activities strategically and effectively. According to Sungur (2007), a significantly positive correlation was found among control of learning beliefs, self-efficacy for learning and performance, task value, and intrinsic goal orientation.

**Method**

**Participants**

One hundred twenty-three students from five high schools in three different states in the U.S. participated in the study, and 113 students completed valid questionnaires. These schools implement a Project Lead the Way (PLTW) curriculum, which supplies middle- and high-school students with hands-on, rigorous, and preliminary courses.
involved in engineering or biomedical sciences. The reason we chose schools that implement a PLTW curriculum is
twofold. First, student participants from those five schools would work on the same projects that have identical
requirements and level of difficulty. Second, the teachers who deliver the classes across those five schools have
received the same training on engineering design activities of PLTW. As a consequence, we assumed that the
variation of student motivation due to instruction could be minimized because of similar instruction for selected
engineering design activities and because results of the study on students’ motivation when engaging in different
engineering design activities are comparable.

These students participated in a Principles of Engineering class that required them to explore technology systems and
manufacturing processes, and address the social and political consequences of technological change through a
combination of activity-, project-, and problem-based learning. These students were required to engage in several
engineering design activities (e.g., marble-sorter design or bridge design).

Measures

This study utilized a modified version of the Motivated Strategies for Learning Questionnaire (MSLQ) survey
instrument. The MSLQ is a self-reporting instrument developed by Pintrich et al. (1991) to assess college students’
motivational orientations and their use of different learning strategies for a college course. Although the MSLQ
was designed for a college course, the researchers chose this instrument for three reasons: (1) This is the only instrument
available that measures motivation with the interest and expectancy components; (2) This instrument has been
widely used in educational research in college and lower-level education courses; (3) Because the course in which
the study participants enrolled (i.e., Principles of Engineering) is college credit equivalent, it was expected that
statements in this survey would be understood by sophomore and junior high-school students.

The modification of the instrument was made by rewording words associated with learning in an engineering design
activity. The intention was to help students focus on their design project as the context of each statement in the
survey. The wording became an essential factor in modifying the instruments because students typically distinguish
between their capabilities for dealing with two or more characteristically different topics or problems within the
same measurement specificity (Bong, 1999). Two open-ended questions were added in the instrument to provide
additional information about the most and least motivating factors.

Validation of the MSLQ and the subscale correlations with final grades were significant, demonstrating predictive
validity. Confirmatory factor analysis tested how closely the input correlations could be reproduced, given the
constraints that specific items fall on. For example, we had six items that were assumed to be indicators of the task
value construct. The confirmatory factor analysis tested how closely the input correlations could be reproduced,
given that those items fall onto one specific factor (i.e., task value). All of the 26 motivation items were tested to see
how well they fit the latent factors. The Cronbach’s alpha coefficients were robust, ranging from .68 to .93 (Pintrich,
Smith, Garcia, & McKeachie, 1993).

Only the motivational scales of the MSLQ (i.e., 26 items) were used in this study. The examples of motivational
scale items are shown in Table 1. The motivational scales included five components. First, the instrument comprises
statements that measure students’ perceptions of why they are engaging in the learning task (i.e., items 1, 16, 22, and
24), an IGO (α = 0.74). Second, statements measure the degree to which students perceive themselves to be
participating in the task for extrinsic reasons (i.e., items 7, 11, 13, and 30), an EGO (α = 0.62). Third, statements
measure students’ perception of how important, useful, and interesting the task is (i.e., items 4, 10, 17, 23, 26, and
27), a TV (α = 0.90). Fourth, statements are present that measure the students’ beliefs that their efforts to learn will
result in positive outcomes (i.e., items 2, 9, 18, and 25), a CLB (α = 0.68). Fifth, statements are present that measure
each student’s expectation to perform the task well and to be self-efficient (i.e., items 5, 6, 12, 15, 20, 21, 29, and
31), a SELP (α = 0.93). Students rated themselves on a seven-point Likert scale, from “not at all true of me” (a score
of 1) to “very true of me,” (a score of 7).

Data collection and analysis

Immediately after completing the design project, students were asked to complete the survey instrument. This was
done to ensure that students had a good recollection of their design experiences. The mean and standard deviation of
students’ interest, expectancy for success, IGO, EGO, TV, CLB, and SELP were calculated. Pearson correlation tests were conducted to find any correlation between the interest and the expectancy for success components. Similar correlation tests were also conducted to evaluate any correlation between each component of the interest and expectancy for success. These correlation tests were conducted to answer research question 1. A simple regression test was used to evaluate the relative importance of contribution of those interest components toward students’ expectancy for success (to answer research question 2). Because some questions in the survey instrument were provided in open-ended format, common themes were categorized into the five motivational constructs to help us further understand students’ motivation qualitatively. Two research assistants helped the researcher categorize students’ responses. Any disagreement in categorization was reviewed and discussed to achieve consensus.

### Table 1. Examples of motivational scale items

<table>
<thead>
<tr>
<th>Motivational scale</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic goal orientation</td>
<td>In a class like this, I prefer course material that really challenges me so I can learn new things.</td>
</tr>
<tr>
<td>Extrinsic goal orientation</td>
<td>Getting a good grade in this class is the most satisfying thing for me right now.</td>
</tr>
<tr>
<td></td>
<td>The most important thing for me right now is improving my overall grade point average, so my main concern in this class is getting a good grade.</td>
</tr>
<tr>
<td>Task value</td>
<td>I think I will be able to use what I learn in this course in other courses.</td>
</tr>
<tr>
<td></td>
<td>It is important for me to learn the course material in this class.</td>
</tr>
<tr>
<td>Control of learning beliefs</td>
<td>If I study in appropriate ways, then I will be able to learn the material in this course.</td>
</tr>
<tr>
<td></td>
<td>It is my own fault if I don’t learn the material in this course.</td>
</tr>
<tr>
<td>Self-efficacy for learning and performance</td>
<td>I believe I will receive an excellent grade in this class.</td>
</tr>
<tr>
<td></td>
<td>I’m certain I can understand the most difficult material presented in the readings for this course.</td>
</tr>
</tbody>
</table>

### Results

#### A. Descriptive statistics of students’ motivation

The descriptive statistics show the mean of students’ interest, expectancy for success, and other motivational aspects (Table 2). A negatively skewed distribution was indicated for all the constructs. Among other motivational constructs, SELP had the highest mean.

### Table 2. Descriptive statistics of students’ motivation ($N = 113$)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest</td>
<td>5.34</td>
<td>1.04</td>
</tr>
<tr>
<td>Expectancy for success</td>
<td>5.46</td>
<td>1.15</td>
</tr>
<tr>
<td>IGO</td>
<td>5.15</td>
<td>1.29</td>
</tr>
<tr>
<td>TV</td>
<td>5.37</td>
<td>1.34</td>
</tr>
<tr>
<td>SELP</td>
<td>5.54</td>
<td>1.29</td>
</tr>
<tr>
<td>EGO</td>
<td>5.50</td>
<td>1.12</td>
</tr>
<tr>
<td>CLB</td>
<td>5.39</td>
<td>1.19</td>
</tr>
</tbody>
</table>

*Note. SD = standard deviation.*

#### B. Correlations between interest and expectancy components

Correlation tests need to be conducted to answer research question 1. As mentioned before, three psychological constructs, IGO, EGO, and TV, will be correlated with the other two psychological constructs, CLB and SELP. Pearson correlation tests indicated a significant correlation between students’ (see Figure 1 and Tables 3 and 4):

1) interest in the design task (Interest) and expectancy for success (Expectancy for Success)
2) self-efficacy for learning performance (SELP) and intrinsic goal orientation (IGO)
3) self-efficacy for learning performance (SELP) and task value (TV)
4) self-efficacy for learning performance (SELP) and extrinsic goal orientation (EGO)
5) control of learning beliefs (CLB) and intrinsic goal orientation (IGO)
6) control of learning beliefs (CLB) and task value (TV)
7) control of learning beliefs (CLB) and extrinsic goal orientation (EGO)

Because the data were collected from two pools of students working on two different design tasks (i.e., bridge and marble-sorter designs), further correlation tests were conducted. The objective was to evaluate if similar significant correlations also existed in students working on the same design task. From the statistical tests, significant correlations also existed between students’ interest in the design task and their expectancy for success. From the correlation tests, it was found that the correlation coefficient ($r$) between IGO and SELP, IGO and CLB, TV and SELP, and TV and CLB were much higher than between EGO and SELP and between EGO and CLB.

Table 3. Correlation between interest and expectancy for success

<table>
<thead>
<tr>
<th>Interest</th>
<th>Pearson Correlation</th>
<th>Sig. (two-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELP</td>
<td>.838**</td>
<td>.000</td>
<td>113</td>
</tr>
<tr>
<td>CLB</td>
<td>.874**</td>
<td>.000</td>
<td>113</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the 0.01 level (two-tailed).

Table 4. Correlations among motivational constructs

<table>
<thead>
<tr>
<th></th>
<th>IGO</th>
<th>TV</th>
<th>EGO</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELP</td>
<td>.804**</td>
<td>.874**</td>
<td>.362**</td>
</tr>
<tr>
<td>CLB</td>
<td>.721**</td>
<td>.704**</td>
<td>.339**</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the 0.01 level (two-tailed).

C. The importance of the interest components in students’ expectancy for success

A simple regression test was conducted to determine the relative importance of the contribution of IGO, EGO, and TV towards students’ expectancy for success. The results revealed that TV ($\beta = .547, p = .000$) and IGO ($\beta = .339, p = .000$) were highly significant predictors of students’ expectancy for success; EGO was not a significant predictor of SELP ($\beta = .050, p = .327$). The three interest-related constructs constituted about 76 percent of students’ expectancy for success (indicated by the adjusted $R^2$ value of .758). From the results, it may be inferred that the motivational
constructs that measure students’ interest (i.e., IGO, EGO, and TV) account for 75.8 percent of the variation in students’ expectancy for success. In fact, there are other factors that might contribute to expectancy for success, such as student’s anxiety and maladaptive behavior, which may include tardiness and unrealistic aspirations.

D. Common themes of students’ motivation

Of the 113 students, a total of 104 responded to the two open-ended questions in the survey instrument. Students were asked to share their thoughts about the most and least motivating factors they experienced during their design activities. Common themes were categorized according to the five motivational constructs used in this study (i.e., EGO, IGO, TV, SELP, and CLB; see Tables 5 and 6). No relevant theme was found for control of learning beliefs.

From the collected themes, it is clear that some themes were perceived as both most and least motivating factors. For instance, a task that was challenging was perceived as the most motivating factor for some students, while it was perceived as the least motivating for others. Similarly, for the time allotted to complete the design task, limited supporting materials available for students to use during design process was perceived differently.

The most motivating themes illustrated how the students responded positively to hands-on experience, task challenge, time challenge, and mastery. Students were also motivated because of their successful performance in solving the design task and interest in the design task itself. The least motivating themes showed that some students were concerned with failure or poor performance, task challenge, and lack of interest. It is clear the students were aware of how the design task was presented to them and how they related the task to their abilities. Difficult or complex design tasks tended to decrease their overall motivation. In other words, unmanageable design tasks may lead students to failure or poor performance.

<table>
<thead>
<tr>
<th>Interest</th>
<th>IGO</th>
<th>EGO</th>
<th>TV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hands-on experience, mastery, task challenge, time challenge, plenty of time available</td>
<td>Successful performance, getting a good grade, comparison and competition, good teamwork, evaluation by others, teacher assistance, supporting materials</td>
<td>Interest in the content or project</td>
</tr>
<tr>
<td>Expectancy for success</td>
<td>CLB</td>
<td>SELP</td>
<td>Ability to master, self-confidence to perform</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interest</th>
<th>IGO</th>
<th>EGO</th>
<th>TV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Task challenge, time challenge, administrative challenge, lack of challenge</td>
<td>Supporting materials, failure or poor performance, bad teamwork, lack of instruction and teacher assistance, getting an unsatisfactory grade, competition</td>
<td>Lack of interest, lack of opportunity to engage in similar project</td>
</tr>
<tr>
<td>Expectancy for success</td>
<td>CLB</td>
<td>SELP</td>
<td>(no relevant theme was found)</td>
</tr>
</tbody>
</table>

Conclusion

The significant correlations between all the three interest constructs (IGO, EGO, and TV) and the two expectancy constructs (CLB and SELP) clearly indicated a strong relationship between students’ interest in design activities and their expectancy for success in completing the design task. These findings conform to modern expectancy-value theories, which suggest that expectancies and values are positively and not inversely related (Eccles & Wigfield, 2002). Strong correlation coefficients between IGO and SELP ($r = .804, p = .000$), IGO and CLB ($r = .721, p = .000$), TV and CLB ($r = .704, p = .000$), and between TV and SELP ($r = .874, p = .000$) imply a more dominant role of intrinsic and task value in the increase of students’ expectancy for success. The results conform to high beta values for IGO ($\beta = .339, p = .000$) and TV ($\beta = .547, p = .000$) in the regression test results. This suggests that students with high self-belief in their effectiveness or confidence are more likely to believe they will perform better in design tasks.
From the expectancy-value model (Eccles & Wigfield, 2002), it is suggested that both expectancies and values lead to performance, persistence, and choice. Students who have low self-perceptions of ability to succeed and control, do not perceive a task to be important, useful, or interesting (low perceptions of task value), may seek to avoid a task rather than to invest effort in completing it. According to Schunk (1991), when students see themselves as effective learners, they are more highly motivated, work harder on learning tasks, expend more effort, and display more self-regulatory behaviors. The level of self-efficacy, therefore, may be related to students’ intrinsic motivation to persist in carrying out a learning task and thereby affects their ability to develop problem-solving skills. One’s expectancy for success may also influence the development of task values.

**Discussions**

The results of this study may have implications in engineering and technology education in grades 9–12. First, in relation to the significant correlation between students’ interest in the design task and expectancy for success, students’ interest in engaging in the design tasks should be positive. A study conducted by Durik and Harackiewicz (2007) suggested that TV plays an important role in students’ interest development. Efforts that promote students’ understanding about the usefulness, interest, and importance of solving the design task are necessary. A teaching approach that emphasizes how a design solution can be used in real-life applications may positively contribute to the increase of students’ task value and interest development. For example, control systems designed by the students in the marble-sorter design project may be used in the detecting system in industry. The level of difficulty and complexity of the task also must be carefully considered before the task is given to the students. Teachers may want to present their students with a design task that is challenging but within their reach. When students perceive the design task to be beyond their abilities, they may become discouraged, which may, in turn, lower their intrinsic motivation.

Second, the 76 percent contribution of IGO and TV toward students’ expectancy for success also indicated other factors that played roles in developing students’ expectation to successfully complete their design tasks. EGO does not seem to be a good predictor for students’ expectancy for success. Smaller correlation coefficients between EGO and SELP ($r = .362, p = .000$) and between EGO and CLB ($r = .339, p = .000$) conforms to insignificant standardized beta coefficient in the regression test ($\beta = .050, p = .327$). Those other factors may be associated with the way the instruction of the design courses was presented. Moreover, in addition to students’ appraisal of challenge and task absorption, issues that relate to students’ self-determination and a feeling of autonomy are also essential elements in intrinsic motivation (Butler, 1987). When students are allowed to make decisions, their interest in learning is likely to increase. When student evaluation is based upon self-referenced information, students develop learning goals. However, when evaluation information is based on normative standards and social comparison (i.e., EGO), students will be more likely to focus on performance goals (Song, 2004), which may increase evaluative pressure and, in turn, may work against intrinsic motivation (Deci & Ryan, 1980). Whenever necessary, teachers may want to carefully exercise extrinsic motivational strategies to increase students’ expectancy for success in a design challenge. Contrary to many studies that found intrinsic and extrinsic motivations to be positively correlated, some studies found otherwise (Wolters, Yu, & Pintrich, 1996).

**Acknowledgments**

This material is based upon work supported by the National Science Foundation under Grant No. 0426421. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**References**


An Investigation into Parent-Child Collaboration in Learning Computer Programming

Janet Mei-Chuen Lin and Shu-Fen Liu

Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan // 1National Lo-Tung Commercial Vocational High School, Yi-Lan County, Taiwan // mjlin@ntnu.edu.tw // liu504@ms22.hinet.net

ABSTRACT

A multi-case study approach was used in this study to investigate how parents and children collaborated with each other when they learned to program in MSWLogo together. Three parent-child pairs were observed directly in a five-day computer camp. Each pair was assigned a total of 33 programming tasks to work on. The observation focused on how parents interacted with their children while they tried to solve the tasks together. Findings from this study indicated that, despite different patterns of interactions exhibited in the three pairs, parent-child collaboration in programming naturally fell into a special form of “pair programming” in which the parent and the child in each pair took the roles of “the reviewer” and “the driver” respectively. As such, children wrote programs in a more systematic and disciplined manner instead of resorting to trial-and-error and tinkering; children spent more time on analysis and design which were considered essential to successful problem solving; the programs they produced were more compact, well-structured, and contained fewer errors. Moreover, children tended to reflect on their solutions more often than if they learned alone. The interviews conducted after the camp revealed that the participating parents and children alike found great pleasure in learning to program collaboratively. In addition, the parents felt that the collaboration enabled them to gain a better understanding of their children and develop a closer parent-child relationship.

Keywords

Parent-child collaboration, Computer programming, Logo, Pair programming

Introduction

The presence of computers in households around the world has changed family dynamics, especially with regard to parent-child relationships. As Papert (1996) has pointed out, the comfort that the younger generations have with computers has made children more independent of their parents in their exploration of the world. He suggested that parents should spend more time trying to find good family computer projects to do together with their children and “use the children’s enthusiasm for computers as a basis for enhancing the family’s learning culture” (p.79). Margolis and Fisher (2002) also emphasized that “parents impart their computer enthusiasm and skills to their children, and through early mastery acquired at home children gain a competence and confidence they carry with them into school” (p.20).

In addition to playing computer games, surfing the net, making a multimedia show, and doing other activities using off-the-shelf application packages, computer programming may be a more educational task that parents may do together with their children. In an attempt to assess the effects of learning computer programming on the cognitive style of 18 1st-graders, Clements & Gullo (1984) concluded that computer programming can increase some aspects of problem-solving ability, including reflectivity, divergent thinking, metacognitive ability, and the ability to describe directions. Programming can also provide parents with a window into a child’s mind because a program has imbedded in it the child’s concepts, strategies, and styles, which could only be the product of the child’s mind (Valente, 1995). When children are engaged in the process of programming, they cycle through the steps of “description-execution-reflection-debugging-description” repeatedly. These steps, particularly the debugging activity, help a child to construct knowledge and learn about problem-solving strategies (Valente, 1995). According to Ellinger (2003), the programming experience is profoundly educational for most people because programming is meticulous; programming teaches self-criticism and responsibility; programming is creative; and programmers communicate, collaborate and share.

Over the past decades many programming languages and environments have been developed that are intuitive and can be easily learned by both children and adults who have no previous programming experience. Notable examples include Alice (http://www.alice.org/), Lego Mindstorms (http://mindstorms.lego.com/), various versions of Logo (http://www.logofoundation.org), Pico Crickets (http://www.picocricket.com/), Scratch (http://www.scratch.mit.edu),
Squeak (http://www.squeak.org/), and Stagecast Creator (http://www.stagecast.com/index.html). Interested readers may refer to Kelleher & Pausch (2005) for a more extensive list of such tools.

The idea that parents can positively influence their children's education is common sense. Practically any teacher will verify that parents who care enough to be involved in their children's learning tend to have children achieve at a higher level. Unfortunately, parent-child collaboration in learning computer skills, especially in learning computer programming, has rarely aroused attention, as evidenced by scarce published studies related to it. Among those very few studies, Armon (1997) found that the LEGO-Logo courses he offered to 18 parent-child (six-grader) pairs helped to foster and cultivate the participants’ thinking skills and creativeness, as well as improve family ties and bring about better understanding between parents and their children; Hughes & Greenhough’s study (1995) showed that seven-year-old children working with an adult performed significantly better than children working without an adult when they learned to use simple Logo commands. More recently, in a set of workshops organized by Hart (2010), which were targeted at girls in grades 4–6 and their parents/guardians, the children experienced pair programming with their adult counterpart while participating in activities such as the Alice storytelling challenge and Pico Cricket design studio. Much of the feedback the researcher received from the parents was positive, leading the researcher to conclude that the workshops succeeded in hitting an age group where it was still fun for students to do activities with their parents.

The research reported in this paper sought to understand how parents and children collaborated with each other when they wrote MSWLogo programs together. Specifically, we tried to reveal the characteristics exhibited in each parent-child pair and across the three pairs during their problem-solving processes. The parents’ and the children’s attitudes toward programming together and the styles of parent-child collaboration were examined.

**Research Method**

This research used a multi-case study approach in which three parent-child pairs were observed directly in a five-day computer camp. Qualitative data were collected through classroom observation, individual interviews and video recordings. In the following we describe the MSWLogo commands that were taught to the parent-child pairs, the profiles of the six participants, and the research procedure.

**The Programming Tool: MSWLogo**

MSWLogo (Microsoft Windows Logo) used in this study is a version of the Logo language (Papert, 1993). It can be downloaded free of charge from http://www.softronix.com/. As a language designed for children, Logo has easy entry routes for nonmathematical beginners, which suits the fourth graders and their non-technology-savvy parents who participated in this research. The set of 14 MSWLogo commands that the participants learned are summarized in Table 1, and the MSWLogo programming environment is shown in Figure 1.

**The Participants**

All students who had just completed the fourth grade at a rural elementary school in northern Taiwan were invited to participate in the MSWLogo camp. Each participating student was required to be accompanied by either of his/her parents for the entire length of the camp. Fifteen parent-child pairs signed up voluntarily, out of which three pairs were randomly chosen for close observation. The reason why only three pairs were chosen for observation was due to the resource-demanding nature of qualitative research. In addition to an instructor who was responsible for teaching MSWLogo and devising learning materials, there had to be one observer deployed for each parent-child pair during the entire period of the 5-day camp. In terms of equipment needs, one digital video camera had to be set up for each parent-child pair so their interactions could be thoroughly recorded. Consequently, only three pairs were chosen for observation. As there happened to be three different parent-child combinations among the 15 pairs, namely, father-son pairs, mother-son pairs, and mother-daughter pairs, we decided to choose one pair from each combination randomly.
Table 1. The MSWLogo commands learned by the participants

<table>
<thead>
<tr>
<th>Command</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>PenUp (or PU)</td>
<td>Lift the pen as the Turtle moves</td>
</tr>
<tr>
<td>PenDown (or PD)</td>
<td>Draw as the Turtle moves</td>
</tr>
<tr>
<td>PenUp + Forward n</td>
<td>Move forward n units</td>
</tr>
<tr>
<td>PenUp + Back n</td>
<td>Move backward n units</td>
</tr>
<tr>
<td>PenDown + Forward n</td>
<td>Move forward and leave a trail.</td>
</tr>
<tr>
<td>PenDown + Back n</td>
<td>Move backward and leave a trail</td>
</tr>
<tr>
<td>Right n (or RT n)</td>
<td>Turn to the right n degrees</td>
</tr>
<tr>
<td>Left n (or LT n)</td>
<td>Turn to the left n degrees</td>
</tr>
<tr>
<td>pensize [pixels]</td>
<td>Change the pen size</td>
</tr>
<tr>
<td>repeat [times] [....]</td>
<td>Repeat a block of commands</td>
</tr>
<tr>
<td>setpencolor [R,G,B]</td>
<td>Change the pen color</td>
</tr>
<tr>
<td>setfillcolor [R,G,B]</td>
<td>Change the fill color</td>
</tr>
<tr>
<td>ClearScreen (or CS)</td>
<td>Erase the screen</td>
</tr>
<tr>
<td>Home</td>
<td>Return to the origin</td>
</tr>
</tbody>
</table>

Figure 1. The MSWLogo programming interface

All the participating children had learned Word, PowerPoint and PhotoImpact in their third- and fourth-grade computer classes before attending this programming camp. In contrast, not all parents had had computer experience. Table 2 shows the three parents’ profiles. As far as the children’s prior geometry knowledge is concerned, they were familiar with basic geometric shapes and could identify key characteristics of those basic shapes, but none of them knew the concept of external angles. MSWLogo was new to all participants, parents and children alike.

The Procedure

The MSWLogo camp was held for 5 days, 3 one-hour periods per day. The instructor was an experienced teacher who had taught elementary computer courses for more than 10 years. In each period the instructor introduced one or more MSWLogo commands and showed a few examples first, he then assigned tasks for the parent-child pairs to work on for the remainder of the period. There were a total of 33 programming tasks assigned during the entire camp. Table 3 shows some of the tasks solved successfully by a parent-child pair.
Table 2. Participating parents’ profiles

<table>
<thead>
<tr>
<th></th>
<th>Case #1 (Father)</th>
<th>Case #2 (Mother)</th>
<th>Case #3 (Mother)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>43</td>
<td>35</td>
<td>39</td>
</tr>
<tr>
<td>Education</td>
<td>junior college</td>
<td>high school</td>
<td>college graduate</td>
</tr>
<tr>
<td>Profession</td>
<td>real estate agent</td>
<td>house keeper</td>
<td>elementary school teacher</td>
</tr>
<tr>
<td>Experience with Software Packages</td>
<td>MS Word, Excel, PowerPoint, Flash</td>
<td>None</td>
<td>Word, Excel, PowerPoint, Flash, PhotoImpact</td>
</tr>
<tr>
<td>Programming Experience</td>
<td>No</td>
<td>No</td>
<td>Yes (COBOL)</td>
</tr>
<tr>
<td>Parent-Child Collaborative Learning Experience</td>
<td>Yes (Chinese calligraphy)</td>
<td>Yes (English)</td>
<td>Yes (Piano)</td>
</tr>
</tbody>
</table>

Table 3. Some of the programming tasks solved by a parent-child pair

<table>
<thead>
<tr>
<th>Task No.</th>
<th>Figure To Be Drawn</th>
<th>MSW/Logo Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>RT 90 FD 100 BK 50 RT 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FD 50 LT 90 FD 50 BK 100</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>RT 90 FD 100 RT 90 FD 200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT 90 FD 100 RT 90 FD 200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT 90 FD 50 RT 90 PU</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FD 50 LT 90 PPT FD 100 LT 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FD 200 LT 90 FD 100 LT 90 FD 100</td>
</tr>
<tr>
<td>3</td>
<td>120°</td>
<td>RT 180 FD 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT 60 FD 100 BK 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LT 60 LT 60 FD 100</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>RT 45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FD 50 RT 90 FD 50 RT 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FD 50 RT 90 FD 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PU FD 50 RT 90 BK 30 PD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FD 100 FD 20 RT 90 FD 130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT 90 FD 120 RT 90 FD 130</td>
</tr>
</tbody>
</table>

The three parent-child pairs were observed the whole time while they collaborated in solving the assigned tasks. The observers used a pre-designed observation sheet to record how a parent and a child collaborated with each other in a programming session. Three observers rotated every day so that each parent-child pair could be observed by different observers. With the consent of the participants, all the programming sessions were videotaped so that transcripts derived from the videotapes could be used to supplement the direct observation data.

On the third day of the camp, the children were given the first hands-on test, in which they had an hour to solve 3 problems. The second test, which contained two problems, was given on the last day of the camp. The children were required to use the more difficult REPEAT command in the programs they wrote for the second test. The test problems can be found in Table 5 near the end of this paper. In both tests parents were not allowed to help their children so that we could assess how well the children had learned.

After completion of the camp, the six participants were interviewed individually. All interviews took place in the classroom where the camp was held. Each interview lasted approximately 30 minutes. When a parent or a child was being interviewed, only the interviewer (i.e., the researcher) and the interviewee were in the classroom. We deliberately separated parents from their children during interviews because we wanted the children to be able to express their true feelings without possible interventions from their parents. The interviews were semi-structured; in
other words, we had prepared a set of questions in advance, but new questions might also be brought up as a result of what the interviewees said. Two specific topics that we wanted to explore were the collaborative problem-solving process and the participants’ interest in collaboration and programming. Table 4 shows the questions for children. The questions for parents were similar except that there might be subject/object interchange for some questions; for example, “How did your parent help you during problem solving?” would be changed into “How did you help your child during problem solving?” and “Did you feel pressure when your parent sat next to you watching you write programs?” would be modified as “Do you think that your child felt pressure when you sat next to him/her watching him/her write programs?”.

Table 4. Interview questions for children

<table>
<thead>
<tr>
<th>I. The Collaborative Problem-Solving Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How did you usually approach a problem? (Hint: Did you usually go by trial and error, or did you always analyze the problem and design the solution before you write the code?)</td>
</tr>
<tr>
<td>2. When you were writing a program, did your parent provide ideas and suggestions often? What did you do when you did not agree with your parent?</td>
</tr>
<tr>
<td>3. What did you usually do to debug a program that did not produce the intended result?</td>
</tr>
<tr>
<td>4. What were the major problems you encountered during problem solving?</td>
</tr>
<tr>
<td>5. In what specific aspects did your parent help you most during problem solving?</td>
</tr>
<tr>
<td>6. Have you ever discussed programming related topics with your parent after class? Have you practiced together at home?</td>
</tr>
<tr>
<td>7. Have you and your parent shared your programming learning experience with other family members?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Interest in Collaboration and Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did you enjoy learning to program with your parent?</td>
</tr>
<tr>
<td>2. Did you feel pressure when your parent sat next to you watching you write programs?</td>
</tr>
<tr>
<td>3. What did you like most and least about learning with your parent?</td>
</tr>
<tr>
<td>4. Will you be interested in learning with your parent(s) again? What in particular would you like to learn with your parent(s)?</td>
</tr>
<tr>
<td>5. What do you think about MSWLogo? (Hints: Is it easy to learn and use? Is it powerful?)</td>
</tr>
<tr>
<td>6. Will you be interested in attending an advanced MSWLogo course or other programming courses again?</td>
</tr>
</tbody>
</table>

Data Analysis

Detailed records of observations kept by the observers for every programming task done by each parent-child pair were edited against video recordings. A combined classroom observation transcript was produced at the end of the camp for each parent-child pair. Conversations between the parent and the child in a pair were divided into topic-related segments and a segment was considered as the unit of analysis for categorization. Content analysis of the observation transcripts focused on two aspects: the problem-solving procedure and collaboration styles. During the first round analysis, one of the researchers marked each unit with a category name (e.g., debugging, question, suggestion, etc.). Results of the first-round analysis were double-checked by the second researcher to ensure thoroughness and objectivity. With respect to the problem-solving procedure, five major categories were derived from the transcript data: (a) problem analysis, (b) design of solution, (c) coding and testing, (d) debugging, and (e) reflection. Regarding collaboration styles, four categories were derived: (a) questions, suggestions and comments, (b) disagreement of ideas, (c) encouragement and praise, and (d) non-problem-solving-related conversation.

An interview transcript was produced for each participant who was interviewed individually. All interview transcripts were edited against audio recordings. The two transcripts for each parent/child pair were combined into one; that is, respective answers for each interview question were put side by side under each question so the researchers could examine if there were discrepancies in the answers given by a parent and his/her child.

Results and Discussion

Findings from this research are presented below in three subsections: collaboration styles and programming processes, attitudes toward programming and parent-child collaboration, and programming test results.
Collaboration Styles and Programming Processes

Case 1: Father + Son (Peter)

Whenever a new task was assigned, Peter’s father always helped Peter to analyze the figure to be drawn. When Peter found the task easy enough, his father allowed him to write the program by himself. Sometimes Peter’s father would offer suggestions, which were mostly ignored by Peter because he liked to try out his own ideas.

Peter was able to solve simple problems correctly and quickly most of the time. When errors occurred in a program, Peter’s father could not offer much help because of unfamiliarity with Logo commands. Therefore, Peter had to depend on himself for debugging, which he did not encounter much difficulty. However, when Peter failed to solve a problem after several unsuccessful attempts, his father would intervene and help him out. Only then did active interactions occur between father and son. An example of their typical interactions is shown below:

(Peter’s father draws a figure on a piece of paper to help Peter to figure out the correct angle measure.)

Father: How much should the Turtle turn to make the figure look good? How about making it 30 here? Would 45 be too big? Now you will need to subtract something from 180 to get a number measure here. How much should it be?

(Peter decided to try 160 degrees.)

Father: Are you sure you want to turn 160 degrees?

(Peter did not reply and did not make any changes. He went ahead to move the Turtle 200 steps forward.)

Father: 200 steps look okay. Now, how much should the Turtle turn here?

(Peter stared at the screen.)

Father: (Pointing at the first angle) Tell me the measure of this angle. How much should it be?

Peter: 140.

(Peter turned the Turtle 140 degrees to the left. The Turtle’s direction did not look right, so he turned the Turtle 20 degrees to the right and moved it 200 steps forward.)

Father: I think you should turn it a little bit more to the right.

(Peter did as his father suggested.)

Father: (Pointing at the screen) So, what’s the measure of this angle finally?

Peter: ...

Case 2: Mother + Son (Sean)

Sean’s mother adopted an authoritative parenting style. Whenever a task was assigned, Sean’s mother would examine the figure carefully, thought up a solution and wrote the program herself, and then she would read the Logo commands line by line for Sean to enter from the keyboard. She was always in charge, busy telling and showing Sean what to do. The following excerpt shows a typical conversation between Sean and his mother:

Mother: Draw the square first.
Sean: I want to draw the triangle first.
Mother: No, you don’t. Draw the square first.
Sean: How far should I move the Turtle?
Mother: 100.
Mother: Now turn 90 degrees to the right. Don’t forget to leave a space between RT and 90. Now FD 100 again, then ... ...

(Sean typed the commands as instructed by his mother.)

Sean: Now what?
Mother: Type REPEAT. R-E-P-E-A-T.
Sean: ... ...

Sean was not allowed by his mother to try anything other than what the instructor asked for, as manifested in the following scenario. Once during class Sean’s mother spent a few minutes talking on the phone outside the computer...
When she returned and found that Sean had added a couple extra squares to a figure, she asked him to change it back to the way shown on the assignment sheet.

_Mother_: What are you doing?! Why did you draw so many tiny squares?
_Sean_: Doesn’t it look better this way?
_Mother_: There should only be two squares. Erase the extra ones!
_Sean_: No, I want these squares to stay.
_Mother_: No, you can’t. Erase them!
_(Sean erased the extra squares.)_

Since Sean was always taking instructions from his mother instead of thinking by himself, he did not know how to correct programming errors and had to rely on his mother for guidance. When a problem was successfully solved, Sean’s mother would review the program with Sean, explaining how the figure had been drawn line-by-line. She would also discuss with Sean about other possible ways to draw the same figure.

**Case 3: Mother + Daughter (Jade)**

When facing a new task, Jade’s mother usually discussed the problem with her first and then encouraged her to suggest possible solutions. Then Jade’s mother would let her try her own way even if it was not exactly what her mother had in mind. When Jade asked for help or paused too long before making any move, her mother would help her analyze the situation and give her some hints, but she never told Jade explicitly what to do without prompting Jade to think and try out her own ideas first.

_(Jade’s mother drew two nested squares on a piece of paper and explained how to put the inner square in the center of the outer square.)_
_(Jade thought for a while and typed in some commands.)_
_Jade_: That looks weird.
_(Jade’s Mother laughed when she saw the figure Jade drew.)_
_Mother_: The inner square should be placed in the center. Draw it again.
_(Jade’s mother watches as Jade redrew the two nested square.)_
_Jade_: (Pointing at the figure she is drawing) If I want to position the square in the center, I should subtract 40 from 120 and divide it by 2, right?
_Mother_: Correct.

... ...
_Jade_: Should it be like this?
_Mother_: No, that doesn’t look right.
_(Jade reached for the tutorial, trying to look for some clues for solving the problem. Her mother took the tutorial away and asked Jade to think hard and find out what went wrong.)_
_Mother_: I think you should change the Turtle’s heading before you draw the line.
_(Jade couldn’t seem to figure out how much the Turtle should turn each time in the REPEAT command.)_
_Jade_: REPEAT is so troublesome!
_Mother_: No, it’s not. Come on, try it one more time.
_Jade_: ... ...

When errors occurred in a program, Jade’s mother usually let Jade find and correct the errors by herself if the errors were obvious, such as a misspelled command or an angle that was too big or too small. However, if the errors looked more serious, Jade’s mother would review the problem with Jade and explain again how to solve it. Then Jade would clear the screen and rewrite the program. If it still did not work, Jade’s mother would write the program herself, try it on the computer to verify its correctness, and then explained the solution to Jade.

Whenever Jade completed a task successfully, her mother would check the figure carefully to see if it looked exactly the same as what was shown on the assignment sheet. If not, she would ask Jade to fix it. After that, they reviewed the solution together and discussed other approaches to do the task.
Summary of Collaboration Styles and Programming Processes

Even though the three parent-child pairs exhibited different patterns of collaboration, the following characteristics can be induced concerning how the three pairs collaborated and proceeded in their problem-solving processes.

- Parents and children partnered up as pair programmers naturally. It is interesting to notice that parent-child collaboration in programming naturally fell into a special form of pair programming, a widely adopted technique in the software engineering world (McDowell, Bullock, Fernald, & Werner, 2002; Williams & Kessler, 2000). When pair programming, two partners sit shoulder to shoulder at one computer; one member is the designated “driver” who controls the keyboard and mouse while actively creating code; the other member is the “reviewer” who constantly reviews the keyed data in order to identify possible errors in the code. With the three parent-child pairs in our study, the children always assumed the role of driver, while their parents were always the reviewer. (This is why we called it a “special form,” because in real pair programming the two partners are required to alternate between the roles of driver and reviewer.) In today’s society, children are often more skilled at computer tasks than their parents due to frequent use of technology. Therefore, parents may not get the hang of MWSLogo commands as quickly as their children. However, parents usually are better equipped with problem-solving skills and geometric knowledge, which enables them to be more qualified as reviewers. Furthermore, it is all too natural for parents to willingly relinquish use of computers to their children so their children can have more chances to practice.

- Analysis and design were emphasized. According to our teaching experience, elementary school students tend to write programs by trial and error. Most of them rush to the computer to key in commands before they really know how to solve a problem. When a program does not produce results as expected, “tinkering” (Perkins & Martin, 1986) is the typical approach to debugging. Consequently, the students may not understand how a program works even if the result turns out to be correct. In contrast, the three children in this study were encouraged (or, rather, required) by their parents to analyze the task and talk about how they intended to solve the problem before they started to key in the code. When errors occurred in a program, the parents usually guided their children through the analysis and design steps again and helped them to identify the mistakes. The proper emphasis on analysis and design led children to write programs in a more systematic and disciplined fashion.

- Programs produced were more compact and well-structured. As a result of the systematic and well-disciplined approach to programming, the programs produced by the three children in this study were more compact and better structured, as compared with those usually generated by children through trial and error when they program alone.

- Reflection on the problem-solving processes was encouraged. Researchers (e.g., Clements & Meredith, 1993) have pointed out that students do not automatically transfer knowledge gained in one situation to another. Therefore, questions that cause students to reflect on what they are doing are instrumental. Unfortunately reflection is usually the most neglected problem-solving step. Students often cannot wait to move on to the next task once they see their programs produce something that looks correct. With the three parent-child pairs in this study, however, it was observed that the parents always took care in ensuring that their children really understood how a problem had been solved. They reviewed the solution with their children and encouraged them to think up alternative solutions.

- The children tended to be overcautious under parental supervision. Under their parents’ supervision, Sean and Jade appeared to be overcautious in making each problem-solving step for fear of making mistakes. For Sean, he might not even have the chance to make mistakes because his mother would have stopped him from going in the wrong direction early on. This seems to be a negative aspect of parent-child collaboration and should be avoided because it deprives the children of the opportunities to explore and to learn from making errors.
Attitudes toward Programming and Collaboration

*Case 1: Father + Son (Peter)*

Peter felt a great sense of achievement in writing MWSLogo programs because he could solve most of the problems without his father’s help. To Peter, his father’s suggestions and guidance were seen rather as interference than facilitation. However, Peter enjoyed learning with his father in general because there were still times when his father could offer useful help.

Prior to this computer camp, Peter’s father had never had the chance to participate in Peter’s classroom activities. He found it a precious experience that they could learn programming together because the camp enabled him to develop closer relationships with Peter and have a better understanding of Peter’s attitude toward learning. He was glad that he could help Peter out a few times. He found Logo easy to learn and use and agreed that programming in Logo could enhance children’s logical thinking skills and increase their understanding of geometry. He said he would like to be able to learn programming with his son again if future camps could be offered on weekends.

*Case 2: Mother + Son (Sean)*

Sean’s mother did not take the camp too seriously at first, but she soon began to feel the time pressure to complete the assigned tasks. Her observation that other teams seemed to be able to finish the tasks quicker brought added pressure on her. Consequently, she spent out-of-class hours learning Logo programming herself and gave Sean extra tutoring at home. Furthermore, for fear that Sean would lag behind other children in getting the tasks done, she decided to think up the solution first and taught it to Sean step by step.

Sean’s mother’s over-anxiousness seemed to have been passed onto Sean. As Sean mentioned in the interview, though he enjoyed writing Logo programs and had fun drawing various figures, he felt anxious when it came to taking tests, and he did not like the pressure. His mother also suggested that tests be excluded from future camps so that children could have all fun and no pressure.

Overall, both Sean and his mother enjoyed learning Logo together. Sean willingly followed his mother’s guidance and felt happy in being able to complete the assigned tasks. He especially liked those tasks that allowed him to draw whatever figures he liked. He said he found great pleasure in playing with figures and was excited when he could create surprisingly interesting ones.

*Case 3: Mother + Daughter (Jade)*

Jade was not very interested in math and science. She also had certain degree of computer anxiety and did not like to use computers. She felt a sense of security when her mother learned together with her; she also liked the way her mother guided her. As an obedient child, Jade always worked hard to meet teachers’ and parents’ expectations. She spent at least half an hour each day at home practicing writing Logo programs in the hope of doing well in the test.

Jade’s mother, who had learned COBOL programming before, liked MSWLogo because she found it easy to learn and considered it suitable for her child’s age. She also appreciated Logo’s immediate, visual feedback that one could verify if one was getting it right. She felt that children could easily gain a sense of achievement through writing Logo programs. She also found parent-child collaborative learning a valuable experience. It not only allowed her to have a better understanding of her child’s strengths and weaknesses in learning but provided her a chance to help her child learn better problem-solving skills.

Summary of Attitudes toward Programming and Collaboration

- Children willingly accepted their parents’ guidance.
  - All the participating parents and children found great pleasure in collaborative learning. The three children willingly accepted their parents’ guidance even though they might not have been allowed enough autonomy as
they had wished (as in Case 2). Moreover, all the three children mentioned that they tended to try their best when their parents were learning with them side-by-side because they did not want to disappoint their parents.

- Parents gained a better understanding of their children. Through parent-child collaboration in programming, the parents gained a better understanding of how their children approached problem-solving challenges and their children’s strengths and weaknesses. They also felt that the course enabled them to develop closer relationships with their children, especially for Peter’s father who rarely participated in his son’s learning.

- All parents and children enjoyed MSWLogo programming greatly. The participants liked MSWLogo because it was easily grasped; it allowed one to draw a relatively complex figure with only a few instructions; it gave immediate, visual feedback; and it increased one’s understanding of geometry. They said they would like to be able to learn advanced features of MSWLogo as well as other programming languages in the future.

### Results of Programming Tests

Table 5 shows the five figures that the three children were asked to draw in the two programming tests, the first three figures for the first test and the remaining two for the second test. Those cells marked with “×” mean that a problem was solved successfully by the child whose name appears in the column head. Out of the five test problems, Peter (Case 1), Sean (Case 2) and Jade (Case 3) solved 3, 2, and 4 problems respectively.

<table>
<thead>
<tr>
<th>Test Problem</th>
<th>Case 1: Father+Son (Peter)</th>
<th>Case 2: Mother+Daughter (Sean)</th>
<th>Case 3: Mother+Son (Jade)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td></td>
<td>×</td>
<td></td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>×</td>
<td></td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>× Problems Completed</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

It can be seen from Table 5 that Jade, though showing greatest anxiety at the beginning of the camp, performed the best. Jade’s good performance may be attributed at least partly to the more proper guidance provided by her mother, who was neither over-involved (as was Sean’s mother) nor under-involved (as was Peter’s father). Another possible reason may be the more effective communication that occurred between Jade and her mother. As compared with Peter’s father whose suggestions rarely generated responses from Peter and Sean’s mother who was constantly giving orders for Sean to follow, the communication between Jade and her mother seemed more effective and productive.

It may not be surprising that Sean did not perform as well as Jade and Peter. As Neuharth (1998) has noted, because authoritarian parents demand more control and dependence from their children, their children tend to be dependent on parental and external validation, which result in weaker decision making skills when they have to solve problems independently. Since Sean’s mother seldom allowed him to contribute problem-solving ideas, it is understandable that Sean might not have learned as much from doing the programming exercises as Peter and Jade did. It was observed that Sean made errors more frequently during the tests. For example, though he solved the fifth problem in
Table 5 successfully, it took four tries. In contrast, Peter and Jade solved the problem after only two tries. As for the fourth problem, he made six attempts to no avail, whereas Peter and Jade were able to solve it quickly.

Conclusion

This paper provides research findings that parents who learned to program MSWLogo together with their children benefited their children’s learning. Specifically, it was found that, despite different patterns of interactions exhibited in the three pairs, parent-child collaboration in programming naturally fell into a special form of “pair programming” in which the parent and the child in each pair took the roles of “the reviewer” and “the driver” respectively. As such, children wrote programs in a more systematic and disciplined manner instead of resorting to trial-and-error and tinkering; children spent more time on analysis and design which were considered essential to successful problem solving; the programs they produced were more compact, well-structured, and contained fewer errors. Moreover, children tended to reflect on their solutions more often than if they learned alone. The interviews conducted after the camp revealed that the participating parents and children alike found great pleasure in learning to program collaboratively. In addition, the parents felt that the collaboration enabled them to gain a better understanding of their children and develop a closer parent-child relationship. These findings are not to be taken as general conclusions though. For the purpose of generalization, a series of quantitative studies are needed to verify each of the above findings. Furthermore, MSWLogo is quite different from most existing programming languages that belong to the “procedural” paradigm. As such, the problem-solving behavior and collaboration styles exhibited may be different if a procedural language was chosen instead.

There have been very few studies that investigated parent-child collaboration in learning. In particular, studies on how children collaborated with their parents in learning computer skills are virtually non-existent. As one of the pioneering studies in this area, the results reported in this paper has attempted to shed a little more light on the potential benefits that parent-child collaboration in learning may bring to children, especially when they try to solve programming problems together. It is hoped that more researchers will be encouraged to delve deeper into how parent-child collaboration in learning may be more effectively implemented both in the classroom setting and outside the classroom. It is also hoped that computer teachers at the K-12 level will be more enthusiastic in involving parents in learning computer programming with their children. Using a properly chosen language, programming can be a great outlet for children to channel their imagination and creativity, and parents can play an active part in it if they choose to.

Acknowledgement

This research has been funded by the National Science Council of Taiwan, the Republic of China, under the grant number NSC 97-2511-S-003-014-MY3.

References


Examination of Co-construction of Knowledge in Videotaped Simulated Instruction

Sitkiye Kuter *, Zehra Altinay Gazi1 and Fahriye Altinay Aksal1

Department of Educational Sciences, Faculty of Education, Eastern Mediterranean University, Northern Cyprus // 1Faculty of Education, Near East University, Northern Cyprus // sitkiye.kuter@emu.edu.tr // zehaltinay@gmail.com // fahaltinay@gmail.com

*Corresponding author

ABSTRACT
The aim of this study was to examine the impact of a videotaped simulated instructional model on prospective language teachers’ co-construction of knowledge with the help of their peers and their supervisor in microteaching sessions. Within the framework of action research, a three-phase collaborative coaching model (TCCM) was developed to encourage trainees’ reflection and instructional awareness regarding teaching and learning processes in videotaped simulated instruction. The implementation of the strategies of qualitative data and method triangulation and collaborative data analysis procedures helped the cross-examination and verification of the data which increased the credibility and objectivity of the research. The analysis of multiple data indicated that the three-phase videotaped simulated instruction is a viable model for trainees’ professional growth. Being involved in video-mediated collaborative teaching and dialogue provided trainees with invaluable opportunities for an in-depth analysis of instructional processes, which raised not only their reflective skills but also professional awareness and development.

Keywords
Microteaching, Videotaped instructional design, Action research, Collaboration, Construction of Knowledge

Introduction

Man is always man-in-the-world and his action cannot be studied adequately if it is separated from his life situation … we must approach and investigate man as integrated being … as a unique person who is embedded in culture, language, society, history and physical world (Lehtovaara, 2001:163).

The realization of professional development using a collective interactive process can be examined within the realm of a constructivist paradigm, which considers learning in the light of situationality. The process of learning ‘how to learn’ and ‘change’ is largely influenced and shaped by the interplay between an individual and contextual variables since it is a shared socially-constructed meaning making process. For the experience to be transformational, one needs to be actively involved in the knowledge creation process (Kohonen, 2001), which requires the incorporation of social negotiation and mediation from multiple perspectives and through dialogue (Santrock, 2001) and learning tools (Borko et al., 2008). Cultural historical activity theory (CHAT) provides a contextual perspective for the significance of collective practical-critical activity because it helps one develop understandings of real world situations, draw meanings from that understanding, and create learning from those meanings (Fox, 2001). The individual could no longer be understood without his or her cultural means; and the society could no longer be understood without the agency of individuals who use and produce artifacts (Engeström, 2001:134). In educational settings, dialogue and activity also play a critical role in promoting learning abilities and facilitating instructional growth through co-construction of knowledge (Postholm, 2008).

In initial teacher education, the means facilitating the development of prospective teachers’ instructional knowledge are considered critical from various dimensions. A review of the literature reveals various studies that emphasize the significance of the incorporation of technology in teacher education with respect to its merits and pitfalls (Goktas et al., 2009; Hixon and So, 2009) and trainees’ creation of knowledge (DiPietro, 2004). It is imperative to integrate current technology in instructional processes (Gulbahar, 2008) because it helps trainees evaluate their gained experiences and learning (Jensen et al., 1994; Kpanja, 2001; Wang and Hartley, 2003) and provides an experience-based and effective learning context (Gomleksiz, 2004). Further, teacher education programs need reform by enhancing the attractiveness of the teaching profession through incorporating a global dimension into teaching practice (Jusuf, 2005). Such integration provides rich situated learning experiences by involving students in social, interactive and active learning processes where they gain skills, strategies and subject-matter knowledge (Ajayi, 2009). In the perspectives offered by CHAT, technology provides active engagement in the learning process, critical
thinking, and communication for the construction of knowledge (Ng’ambi and Johnston, 2006). Video - as a tool for fostering productive discussions and negotiation among trainees within microteaching - provides a supportive, critical, and evaluative environment in enhancing professional development (Glazer et al., 2005; Borko et al., 2008).

Microteaching, as a simulated training technique, has been used for various stages of trainees’ growth. Wallace (1991) considers microteaching as a deep-rooted and highly valuable teacher education technique; however, he also states that ‘...nothing works so well as using real learners. Only then is the true nature of the teaching and learning process revealed at the appropriate level’ (p.101). In this respect, the extent of the effectiveness of microteaching sessions on trainees’ growth in internship can be debatable because simulated classroom discourse may not mirror the actual teaching and learning context; yet, such environments – when supported with technology in a collaborative and reflective learning context – can help construct knowledge.

Another dimension of growth is through reflective dialogue, which is regarded as an inseparable part of a meaning making process. A reflective process, theoretically, may lead to professional growth through a developmental process involving a critical assessment of the existing culture of practice, which is referred to as internalization (Edwards, 2007). It can be conceptualized that self-reflection arises ‘through internalizing the perspective that the other has upon self, followed by self taking the perspective of other upon self’ (Gillespie, 2007:682). Facilitating reflection, interaction and collaboration is critical since multiple perspectives nurture and shape knowledge production by linking the intra-psychological dynamics of reflection to the inter-psychological - social psychological - dynamics of collaboration (Cornish et al., 2007) and professional development (Gazi A., 2009; Jonassen, 1991). According to Wang and Hartley (2003), video technology is effective in assisting trainees’ to develop professional knowledge and dispositions. However, ‘the influence of different ways in which video technology is used on what pre-service teachers learned and how they learned’ has not received considerable attention (p. 130).

Lack of actual teaching experiences has always inhibited trainees from gaining optimum benefit in their professional development during their practicum experiences. Therefore, a three-phase collaborative coaching model (TCCM) was conceptualized. The model incorporates collaborative and reflective dialogues among trainees and their supervisor to encourage trainees’ reflection and instructional awareness as regards their instructional processes based on video-integrated microteaching sessions. The primary aim of this conceptualization, and the research reported here, was to examine the impact of the videotaped simulated instructional model on prospective language teachers in co-constructing knowledge in microteaching sessions. Considering the significance of socially-constructed knowledge for development, as stated in CHAT, this study is significant because, using an action research model, it attempted to critically scrutinize how this holistic model can be applied to teacher education. This was accomplished by depicting, in a broad spectrum of qualitative evidence, how videotaped stimulated instruction in microteaching in English language teacher education was organized and implemented. In this regard, it may provide invaluable insights concerning the effectiveness of the model in preparing trainees for real school experiences.

**Methodology**

The overall framework of the research design for the study was based upon theoretical and methodological premises that support a conception of learning as a socially constructed interpretation of reality (Denzin and Lincoln, 2003). The exploration of the interactions among individuals (Creswell, 2003) helps capture the complexities of the phenomenon through the participants’ perceptions and experiences.

**The Three-phase Collaborative Coaching Model**

The TCCM was designed around three developmental sessions (during seven weeks) to gradually prepare and equip trainees with the thinking skills and professional dispositions, deemed necessary, to be effective beginning teachers when they start their actual teaching experiences. Within the conceptual framework of constructivist teaching and learning processes, and with the support of technology, the practicum was designed on the premises that learning is an interactive, dialectical, collaborative, negotiative and reflective process. Within this framework, trainees were expected to collaborate and negotiate with peers to create pedagogical knowledge for their future professional learning and growth. All microteaching classes used a repeated cyclical framework (see Figure 1) in three phases, starting from pre-sessions and continuing during the while- and post-working sessions. These three phases provided
not only the means for trainees’ reflective analysis of the microteaching sessions but also the grounds for collecting data during the second step of the action research cycle (Figure 2). The phases of the TCCM (pre-working session, while-working session, and post-working session), that mirror the second step of the learning cycle in Figure 2, are described more fully below.

Figure 1. The model of collaborative videotaped simulated microteaching

**Pre-working Session:** Though not a part of the trainees’ video-taped instructional presentation, this session aimed to equip the trainees with some pertinent foundational theoretical knowledge related to instructional issues and to prepare each group of trainees for microteaching presentations that would be videotaped. To achieve this, each group was initially asked to prepare a lesson plan and instructional materials for a teaching point, either a language skill or content area, which had been identified at the beginning of the program by each group (seven in total). The groups were composed of one, two or three peers. Before their presentation, trainees were assigned to read materials related to language skills or content areas. Also, they were expected to revise and improve the quality of their lesson plans and materials by liaising with the instructor prior to their presentation. This pre-conferencing session aimed to raise trainees’ awareness of their lesson plans.

**While-working Session (50 minutes):** This session aimed to involve each group of trainees in peer-teaching presentation which was videotaped to be viewed and discussed in post-working session. Each presentation was timed for thirty minutes and followed by presenters’ self-evaluation for twenty minutes in class. At the beginning of the session, all presenters and peers were given an appraisal form including the aspects of lesson planning skills, teacher personality and talk, warm-up, presentation, planning, feedback, and classroom management. The presenters also distributed the lesson plan to their peers to facilitate their peers’ observation and evaluation of the presentation. They also informed their peers about the level of students that they were expected to address during the presentation. Peer-presentation was conducted and simultaneously videotaped. Meanwhile, the presenters were appraised by their peers in the class. Each peer wrote comments indicating the strong and weak aspects of the presentation using an appraisal form. This session ended with the presenters’ oral self-evaluation.

**Post-working Session (100 minutes):** The ultimate aim of this session was to raise trainees’ awareness of instructional processes by involving them in collaborative discussions after each presentation while viewing the videotape. By stopping the videotape at certain intervals, all trainees discussed and reflected on the presenters’ teaching performance and instructional stages by considering the aspects in the appraisal form. Meanwhile, the teacher, by linking the practical processes to theoretical basis, guided and facilitated co-evaluation process. The session ended by focusing on the aspects that needed further development and the reasons why. After the lesson, all trainees, including presenters, were required to write a reflective journal by focusing on the strengths and weaknesses of the lesson and suggesting ways for improvement and further development. They were also expected to write what they learned from the microteaching and collaborative evaluation. Besides, the presenters were required to evaluate their own performance as regards its strengths, weaknesses and make suggestions for the improvement of weak areas by writing a self-reflective report. Both reports were to be submitted to the instructor before the following microteaching session.
Research Design and Approach

In educational professional practice, understanding how to effectively interpret people’s intentions, capacities and responses during learning, which provides insights about what they are experiencing, is crucial to successful teaching. Therefore, a qualitative research design, which employs a wide range of interconnected interpretive practices, was chosen as an appropriate research design in this research study (Denzin and Lincoln, 2003).

Qualitative research includes naturalistic, inner perspective, emic, and soft data, and encompasses the key concepts of meaning, common-sense understanding, definition of situation, social construction within an inductive process (Bogdan and Biklen, 1998). It is a reflective process, whereby researchers try to understand how meaning and experiences are constructed (Creswell, 2003). Thus, a qualitative research design was deemed most appropriate for the exploration of prospective language teachers’ co-construction of knowledge with their peers and supervisor in microteaching sessions.

Action research was employed to systematically examine a problem within a specific context and seek appropriate solutions based on induced change of professional practices and a learning cycle rationale. Action research, as an evaluative tool, provides an environment that improves rationality and justification of professional practices within a self-reflective context, mediated through collaborative activities, group support and assessments. So, the steps of the action research cycle (Figure 2) were adopted and implemented to examine the impact of collaborative coaching in videotaped, simulated instruction on trainees’ development with respect to a constructivist learning and teaching process (Johnson, 2002; Mills, 2003). In the second step of the cycle, the implementation of the three phases of the TCCM (Figure 1) generated multiple qualitative data which reflected trainees’ views on their professional growth. This repeated cyclical framework supported the trainees’ presentations in microteaching sessions, their collaboration, reflection and evaluation of the instructional and learning processes incorporated in the model. In the following steps of the cycle, the analysis and evaluation of multiple data drawn from the TCCM provided new understandings as to the areas the trainees displayed professional development. The final two steps of the cycle led to further suggestions as regards the model.

![Figure 2. Action research cycle (Mills, 2003)](image)

Context and Participants

The participants of the study included all sixteen interns enrolled in the practicum, a fourth-year second semester course, which aimed at improving trainees’ teaching performance and developing an awareness of how students learn by involving the interns in reflective teaching practice. Seven microteaching lessons - one per week - were conducted throughout seven weeks. The trainees voluntarily participated in the study and the data were collected through purposive sampling in order for an in-depth examination of the learning processes as they unfolded.
Data Collection Techniques and Analysis

Multiple qualitative data collection instruments - interviews, self-reflection reports for micro-teaching sessions, and reflective journals - were used for the in-depth examination of the issue under investigation (Denzin and Lincoln, 2003). Ensuring rigor is critical during data collection and analysis processes of the research because, without it, the research will lose its trustworthiness and utility (Morse et al., 2002). Content - thematic - analysis was employed to interpret all qualitative data by considering the key themes in relation to the research focus and questions (Altinay and Paraskevas, 2008). The implementation of the strategies of qualitative data collection, method triangulation and collaborative data analysis procedures helped the cross-examination and verification of the data from all sources. This increased the credibility and objectivity of the research. The following series of data analysis phases were employed in the study: Organizing raw data for transcriptions, transcribing data around pre-determined and emerging themes and codes, member checking, creating matrices for each theme, coding the data on matrices, inquiry audit confirmation of the data by the participant researchers, and verifying data through data triangulation (Kuter and Koç, 2009). For ethical considerations, confidentiality of the data and anonymity of the participants were ensured during data analysis processes by using coded numbers for all participants. The subjects were not compelled to participate in the study, but were invited to be a part of the study. Before videotaping trainees’ presentations, their consent was obtained. They voluntarily agreed to be videotaped and requested that they receive a copy of the tape to keep as reminiscence.

Findings and Discussion

The in-depth examination of the triangulated data yielded invaluable findings regarding the impact of the collaborative simulated instruction on the trainees’ professional growth in the practicum. Based on this evidence, the three-phase collaborative microteaching was deemed as a viable model having a notable influence on the trainees’ construction of professional knowledge as far as the following facets are considered.

Video as a Means for Professional Development

The data collected from reflective journals and interviews exhibited a promising parity concerning the importance and necessity of the incorporation of video technology during while- and post-working sessions. Almost all trainees considered video as an indispensable part of simulated discourse in facilitating the transformation of their conceptions of teaching and learning from the following aspects.

First, video-taping sessions heightened trainees’ self-awareness of the professional attributes they were expected to develop. Almost half of the trainees indicated increased awareness about certain teaching points which they did not perceive during presentations. ST2 stated, “When I couldn’t concentrate on my peers’ presentations, I found the opportunity to give more consideration for those points during video watching. I became aware of such points that I couldn’t notice during my peer’s presentation.” Further, ST5 and ST13 reported their heightened awareness of the teaching aspects they needed to develop as a result of ‘focused attention’. That watching video broadened trainees’ consciousness towards their posture, use of language and voice, and pronunciation mistakes in English was underlined by some trainees. One of the trainees reported that, when watching her video, she noticed her excitement. Stressing the significance of video watching, ST8 remarked, “On the spot evaluation is difficult. Both watching and evaluation facilitated in-depth evaluation of our faults.”

Second, the data demonstrated that nearly half of the trainees considered video as an invaluable tool facilitating trainees’ reflection on practice. These trainees also noted that the evaluation of presentations via videotaping facilitated their involvement and development of critical reflection. ST15 underlined, “I had the chance to watch with a critical eye. During the presentation I do not have so much time to think about the lesson.” According to these trainees, the trainees’ observing their peers using an appraisal form and a lesson plan during while-working session facilitated their evaluation during video watching in the post-working session since they developed awareness towards the points to be evaluated. ST5 reported, “Both the appraisal form and the lesson plan helped me to be involved in deep reflection. Watching and reflection should go hand in hand after presentation and be followed by writing our reflective journals. All these complement one another.” ST6 also supported this by saying “evaluating peers’ presentations via video helped me to critically evaluate peers from a wider perspective.” Next, almost all
Trainees reported that video-taping provided the grounds for the exploration of teachers’ personal and professional traits from multiple angles. More than half of the trainees stressed that, by watching their peers, they actually made inferences about the things that were not related to teaching skills and processes. Watching and re-watching was deemed useful since, rather than relying on one idea, it helped trainees to put themselves in other’s shoes and empathize with them (ST12). As to ST6, watching video, minute by minute, widened trainees’ awareness and perspectives and contributed to their growth.

Considering the difficulty of being involved in instantaneous self-evaluation in practice, video can be considered as a real third eye which can provide the proof of trainees’ actual teaching performance in their presence and give them a chance to be involved in self- and peer-evaluation and see their strengths and weaknesses more openly. In this regard, audio-visual technology can play a significant role in providing trainees with more chances to see and notice instructional behaviour.

Thus, reflection at technical, contextual and dialectical levels (Taggart and Wilson, 2005) is an inseparable part for instructional development. When video technology is integrated in this technical process as a means for reflection, each session might gain meaningfulness in terms of the inferences drawn. ST6 emphasized, “I learned that whatever the reason is I should focus on my posture because when I see myself on the camera I see a very passive, weak person with low esteem, so students can use this.” Effective technology integration requires teachers to obtain learning experiences with the context of their teaching so they can practice, reflect and modify their practices (Glazer et al., 2005:57). It also provides trainees with rich and diverse instructional experiences to observe and reflect for the intention of acquisition and development of pedagogical content knowledge (Wang and Hartley, 2003).

Borko and his friends (2008) support the integration of video as a professional development tool when participants are involved in reflective and creative conversations within a group. Watching and reflecting on the same video in collaborative communities helps trainees develop a collective expertise from multiple angles (Taggart and Wilson, 2005). To this end, video technology is an effective observational tool, which can foster trainees’ accurate self-evaluation (Anderson and Speck, 1998), reflection within the context of teaching (Jensen et al., 1994), and trainees’ pedagogical knowledge (Kpanja, 2001).

**Collaboration as a Process for the Construction of Knowledge**

The in-depth examination of multiple data exhibited that the phases of TCCM provided invaluable grounds for trainees’ development relative to several dimensions.

Before discussing trainees’ growth, it is crucial to underline their weaknesses in instructional processes. In self-reflective reports, almost all trainees were found to have problems with teacher talk and lesson planning. The trainees’ problems with their use of voice, language, and adjusting their pace of speech were highlighted by eight, five and four trainees, respectively. Regarding planning, while six trainees reported their weaknesses in preparing accurate materials for the implementation of tasks and proper visuals to sustain attention, two trainees underlined their incapability in planning learner-centred tasks. Concerning teacher personality, the problems highlighted were lack of confidence (5 trainees) and enthusiasm (3 trainees), nervousness, excitement (2 trainees), and lack of creativity (1 trainee). Besides, a few trainees experienced difficulty in motivating students in warm-up sessions, and giving clear and purposeful instructions and specific feedback. Finally, posture and time management problems were reported by two trainees, respectively.

The data displayed that videotaped presentations and collaborative discussions promoted trainees’ awareness and growth in various areas of teaching. Regarding teacher talk, almost all trainees highlighted their improvement in using their voice. While six trainees reported their awareness of inaccurate use of English language, four trainees reported improvement in raising and making their voice audible to class. Only one trainee underlined his awareness of pronunciation problems in English. Concerning planning of the lesson, more than half of the trainees stressed their growth in planning various materials for the lesson. The trainees’ progress in planning learner-centred tasks and using the board in a more organized way was mentioned by four and eight trainees, respectively. While nearly half of the trainees emphasized their growth in preparing accurate hand-made materials, some trainees reported their consciousness of integrating visual materials for sustaining attention. The importance of a contingency plan, integrating various technologies and teaching techniques, listening tasks and music, and various sitting arrangement
organizations was highlighted by one trainee. Regarding warm-up, more than half of the trainees stated that they got the importance of motivating students in warm-up. As to feedback, almost all trainees reported the areas they developed themselves in giving feedback. Almost half of the trainees stressed their growth in giving immediate feedback and simpler and clearer instructions, respectively. While six trainees developed awareness of checking students’ comprehension, three trainees developed better skills in using praise and self- and peer-correction. Considering teacher personality, the following attributes of friendlier attitude, confidence, and enthusiasm was reported by nine, five and four trainees, respectively. Few trainees reported that they became more well-organized and creative as a result of microteaching processes. Regarding classroom management, nearly half of the trainees developed awareness about the importance of classroom and time management, respectively. Four trainees also reported their progress in controlling disruptive behaviour. Although trainees reported no weaknesses concerning their lesson planning skills before the video-taping experiences, they stressed their progress in writing a lesson plan. Almost all trainees focused upon the importance of a well-organized and well-thought-out lesson plan. Nearly half of the trainees became aware of how critical it is to consider the preparation of a plan for using the board during planning. Nearly half of the trainees stressed their progress in writing performance objectives and integrating various activities and visuals in a lesson plan, respectively. Finally, more than half of the trainees stressed their development in contextualizing teaching points after the warm-up. As to ST11, “I didn’t go to school as an empty tin. In microteachings, I reached certain awareness about how to use board and developed myself in contextualization of my presentation.” Five trainees also reported how critical the theoretical aspects they studied before are. ST1 noted, “We studied theory, yet without its application everything is in the air. Viewing with examples imprinted everything on our mind.” Considering the overall effectiveness of microteachings, ST7 said, “Observing and discussing the lesson with different nearly twenty eyes, writing reflective reports all contributed to my professional growth and my teachings ... These sessions laid strong foundation on our professional growth and teaching experiences.” ST4 remarked, “My cooperating teacher at school couldn’t find any weakness in my teaching. She only said ‘your language’ because we went there having been already equipped with necessary teaching skills.”

The data drawn from self-reflective reports and interviews showed that trainees found the processes during while-and post-working sessions significant from several aspects. Firstly, observing peers with an appraisal form and lesson plan helped five trainees to be reflective and three trainees to be objective during observation. The form was deemed to broaden trainees’ awareness towards the areas they need to focus by eight trainees. Having been given both the appraisal form and lesson plan was considered effective by five trainees, because having both helped them to go into instructional details, observe peers from wider angles, and follow the presentation in the light of objectives. ST13 reported, “I learned how to write a lesson plan by observing through my peer’s lesson plans. This contributed me a lot, decreased my excitement and increased my confidence.” Observing peers with an appraisal form and lesson plan helped ST6 and ST12 to put themselves in the shoes of both a teacher and student and gain more awareness about instructional processes.

Second, involvement in collaborative discussions helped some trainees to gain awareness regarding their instructional processes. ST13 underlined, “Without collaboration, one cannot develop. I got aware of how to deal with my bad aspects. I liked ‘learning by doing’ and learned that students learn when they are active and the teacher talks less.” Besides, the contribution of collaborative dialogues on trainees’ development was reported by some trainees. As to ST5, “Our friends were giving the instructions after they delivered the material. This should have been done before in order to draw students’ attention to help them to be aware of what to do. This also happened to me.” She also added that when a lesson was supported by visuals and creative activities then it became efficient and enjoyable for everyone and she learned the use of good illustrations in order to contextualize the teaching point. Next, being engaged in reflective dialogues from multiple angles - self- and peer-evaluation, and teacher evaluation - was deemed as one of the most significant phases of videotaped simulated instruction since it helped all trainees to gain a deeper understanding of their development. All trainees considered those dialogues, under the guidance of the critical peer-mediated feedback, through TCCM to be highly beneficial. While four trainees reported that they gained by evaluating peers and looking into matters from multiple perspectives, two trainees found this phase vital since they mixed both theory and practice and reflected on instruction. Both ST9 and ST11 reported that, though microteaching is a fake environment, they learned reflecting on their peers. ST7 stated, “A house without a foundation is impossible. The stronger the foundation is, the stronger the house is. The feedback given was the cement and strengthened our house.” This showed that the feedback given from multiple perspectives raised trainees’ awareness regarding the aspects that one could not notice when involved in instructional processes. The facilitative role of the teacher in providing guidance and supervision was also reported as a key element of this process.
Finally, writing reflective journals in post-working session was reported to help trainees’ growth. Almost all trainees reported that writing reflections helped them to learn how to evaluate peers, which helped them to evaluate themselves as well. ST5 reported that she delivered more effective teaching by noticing her mistakes in reflective writing. As to ST3, “I developed by evaluating myself. I always thought about my peers and my own weaknesses. I tried to find ways out and employ those ways in my actual teachings.” Within collaborative evaluation, particularly teacher feedback was underlined as one of the most critical variables necessary for the realization of collaborative dialogues in post-working sessions by all trainees. Ten trainees considered teacher feedback critical for the collaborative evaluation process. Three trainees reported that they wouldn’t gain from the discussions without teacher feedback. ST8 noted, “Like we underline sentences, important sections, in a book, the teacher highlighted the parts we need to be careful.” The effectiveness and significance of immediate feedback was highlighted by ST7. She said, “You didn’t give us ready-made thought but provided us with the foundations of how to think. Think about a piece of dough. You can shape it when it is warm, but when it gets cold, you cannot work on it.”

The triangulated data demonstrated that being involved in collaborative dialogue during TCCM on one hand helped trainees improve their weaknesses in planning a lesson with respect to materials preparation, teacher’s feedback, use of voice, warm-up, and teacher personality, on the other hand assisted trainees to gain awareness and develop themselves concerning instructional aspects that they were not aware of at the beginning of microteachings. Increased awareness helped all participants to be involved in collaborative, reflective, and critical dialogues which increased the value of those sessions in terms of their professional development (Chalies et al., 2008). Collegiality in the form of collaborative reflective dialogues in microteaching helped trainees interrogate their personal theories and constructs and the taken-for-granted (Francis, 1997) and gain awareness and progress due to the incorporation of self- and peer-evaluation and teacher evaluation. In this regard, involvement of a master teacher in this process was considered critical (Roth et al., 1999). Further, video was considered a significant and complementary element facilitating collaborative reflective dialogue and trainees’ instructional awareness and growth. Video-tape supported microteaching was reported to provide concrete actual experiences ‘to bridge the gap between the abstract and real’ (Brent et al., 1996). Within this framework, negotiation of feedback under the guidance of the instructor and the theories the trainees studied before assisted the collaborative reflective process and helped alleviation of trainees’ growth. Thus, conceptualizing TCCM on the basis of a three-phase video-integrated collaborative reflective process heightened trainees’ awareness and self-reflection which facilitated their growth as regards the multifaceted reality of teaching processes. This supports the idea that learning to teach is a complex phenomenon which is built upon experiences gained in social contexts (Freeman and Johnson, 1998) and helps each prospective teacher’s perceptions and beliefs to develop in complex interaction with contextual experiences (Aksal A., 2009).

**Conclusion and Recommendations**

This study reported and discussed the findings of a three-phase collaborative video-integrated simulated instruction model, which yielded positive results on raising trainees’ awareness, reflection and growth in pedagogical processes. The findings exhibited that all phases of TCCM provided the instructional means for the trainees to bridge theory and practice within a collaborative, negotiative (Kaasila and Lauriala, 2010), and reflective learning environment, all of which facilitated their professional awareness and development. Engaging students in a constructivist-based learning environment encourages their reflection, communication and collaboration skills, thus facilitating their development (Neo and Neo, 2009). This study showed that examination of personal theories in collaboration with a supervisor and the integration of critical friendship are the aspects that need to be considered in microteaching as well as integrating video technology (Francis, 1997). Although microteaching in teacher education has been considered as an artificial (Bell, 2007) and simplified instructional setting (Fernandez, 2010), these sessions, when supported with technology, provided the means for promoting professional knowledge (Borko et al., 2008; Neo and Neo, 2009), consciousness of new possibilities for professional growth (Zellermayer and Ronn, 1999), critical thinking and reflection (Neo and Neo, 2009), and collaborative learning (Borko et al., 2008). Francis (1997) also supports that microteaching helps to develop critical exploration of technical and theoretical aspects. Moreover, it enhances trainees’ self-evaluative awareness and competence building. Video, as a cognitive tool, is considered critical for microteaching and professional development (Borko et al., 2008; Osam and Balbay, 2004). Besides the aforementioned cognitive processes, the findings exhibited that alteration of trainees’ competencies takes place through modelling (Francis, 1997), involving trainees in social processes of professional development. Trainees might imitate and reproduce the behaviour they liked by attending to someone else’s performance (Cruikshank et al., 1995). Such social learning may foster personal and professional development. Further, collaborative discourse
provides the means for professional development through scaffolding and negotiation of feedback. Rather than an individual initiative, a collaborative undertaking - a partnership - was constructed (Kuter and Koç, 2009). Therefore, supporting development through socially constructed knowledge is another dimension crucial for growth. Without being actively involved in various social contexts and processes, one cannot encounter situations and cases from which certain meaning can be generated. Thus, the social dimension of human being gains meaningfulness when actively being involved in social processes, which are not only limited to interactions but the continuing and multifaceted interplay between social processes and mind for the (re)construction of knowledge through internalization (John-Steiner and Mahn, 1996). At this stage, technology plays a critical role since its integration provides cooperative and collaborative (Hwang et al., 2008) and self-directed learning environments (Kayler and Weller, 2007) which enables learners to be at the centre of instruction, construct knowledge through multiple perspective and active learning strategies (Karagiorgi and Symeou, 2005). As Dipietro (2004) highlights,

engaging pre-service teachers in the constructive processes of analyzing, adapting, testing, negotiating, retrying, and reflecting allows them to examine teaching and learning and to determine for themselves how to teach, how students learn, and how technology can support learning. (p. 73)

Kpanja (2001) also stresses that videotaped instruction helps prospective teachers display more significant progress in the mastery of teaching skills when compared with the ones not involved in videotape-supported microteaching. Although microteaching cannot be replaced with other kinds of instructional experiences (Wallace, 1991), this study exemplifies how a model of videotaped simulated instruction, as a whole, can equip trainees with the necessary personal and instructional skills before being involved in actual teaching experiences at schools. In this respect, the findings of this study may help program designers in initial teacher education to construct video-supported collaborative microteaching learning environments to enhance reflection and professional growth. Collaborative professional approaches in technology supported environment can assist to overcome certain hands-on barriers like ‘lack of on-site support and lack of authentic learning experience’ and, therefore, enhance development (Glazer et al., 2005:65). Although the study yielded highly positive results as regards the TCCM, longitudinal and qualitative studies (Wang and Hartley, 2003) need to be conducted to determine its effectiveness on trainees’ growth over a longer period of time. If rigorously conducted, the research agenda outlined also might provide more insights into the value of the model and its limitations.

References


Prospective EFL Teachers’ Perceptions of ICT Integration: A Study of Distance Higher Education in Turkey

Murat Hismanoglu

Akdeniz University, Faculty of Education, Department of Foreign Languages Education, Division of English Language Teaching, Antalya, Turkey // hismanoglu@gmail.com

ABSTRACT

This paper aims to investigate the perceptions of prospective EFL teachers in the distance higher education system toward ICT implementation in teaching English as a foreign language. The majority of respondents who expressed negative attitudes to ICT integration found the nature, level and delivery of the training inadequate and accordingly affirmed that they do not feel sufficiently competent to use ICT in their future subject teaching without having sufficient prior knowledge of ICT dissemination. The results of the study imply that training that will enable teachers to become competent in and receptive to ICT is quite critical in distance education realms.

Keywords

Prospective EFL teachers, Distance higher education, ICT integration, Language instruction

Introduction

Information and Communication Technology (ICT) has been promoted as a platform for providing learners with opportunities in any field (King, 2002; Rovai, 2002) and many significant studies have been conducted on the integration of ICT into classroom teaching to complement and modify the pedagogical practice (Hennessy et al., 2005). Due to ICT related development in the field of education, countries regard ICT as a potential tool for change and innovation in education and so make investments in ICT integration (Eurydice, 2001; Papanastasiou and Angeli, 2008). For instance, Turkey spent 400 dollars per person and allocated eleven point seven percent (11.7 %) of its budget to ICT. This rate is lower than the rates of Europe and Central Asia, since they allocate twenty two percent (22 %) of their budget to ICT implementation, yet it is still higher than the rates in developing countries (World Bank, 2007). Therefore, the Ministry of Education in Turkey attempted to set up computer laboratories and provide Internet connection in schools. To illustrate, the rate of schools with Internet connection increased from forty percent (40 %) (World Bank, 2007) to sixty-eight percent (68 %) (SPO, 2008) and it is projected that this rate will have increased to ninety-six percent (96%) by 2011 (SPO, 2006).

Although investments in ICT for educational innovations and improvements purposes have been continuing, the needs of teachers who will employ it in the classroom as a staple part of the curriculum is disregarded (Niederhauser and Stoddart, 2001; Vacc and Bright, 1999). ICT does not have an educational value in itself, but it becomes precious when teachers use it in the learning and teaching process. As Shakeshaft (1999, p. 4) notes, ‘just because ICT is present does not mean that students are using it’. The impact of ICT is strongest when used in a particular content area and further supported by use across the curriculum (Ward and Parr, 2010). Since teachers are the key figures to utilize ICT in educational settings productively and to help integrate ICT into the curriculum, they need support and training to disseminate ICT integration into their classrooms.

Roblyer (2002) found that many pre-service teachers are still entering universities with little knowledge of computers and appropriate skills as well as lacking positive attitudes toward ICT use in the classroom. Moreover, Gunter (2001) states that many higher education institutions are still failing to prepare pre-service teachers for positive technological experiences. Hence, it is unlikely that teachers will be able to transfer their ICT skills to their students and encourage them to implement ICT when they themselves have negative perceptions toward ICT deployment (Yildirim, 2000).

As highlighted by a variety of substantial studies, however, not all teachers have been willing to introduce ICT into their language classrooms. In the last decade, a steady stream of work has variously addressed this issue (e.g., Mumtaz, 2000; Williams et al., 2000; Galanouli and McNair, 2001; Baylor and Ritchie, 2002). Studies have also shown that, for the younger generation of teachers, the basis of this unwillingness is sometimes to be found in the training on the use of ICT provided in the teaching and learning process (e.g., Watson, 1997; Murphy and Greenwood, 1998; Strudler & Wetzel, 1999).
As Rogers (1995) postulates in his theory of *Diffusion of Innovation*, technology adopters’ perceptions are indispensable to the innovation-decision process. He suggests that studies should focus on users’ attitudes toward ICT integration in the early stages of technology implementation. Perceptions (or beliefs or intentions) are being considered the cognitive components of attitudes and the literature shows that pre-service teachers’ perceptions influence intentions which in turn influence behaviour (Ma et al., 2005; Dillon and Gayford, 1997). For instance, according to the model proposed by Ma et al. (2005), the teacher’s intention to use ICT can be predicted by his/her subjective perceptions of its usefulness. As a general rule, and to expound the relationship of perceptions-intentions-behaviour, it can be stressed that the more favorable perceptions are, the stronger the intentions to perform the behaviour in question will be (Ajzen, 2007). These strong links were also investigated by researchers in many other studies (Norton et al., 2000; Hu et al., 2003) and a high degree of overlap was found between the scores measuring participants’ perceptions and those measuring their intentions and the positive beliefs and intentions expressed by both scores indicated that the participants were expected to carry out their intentions.

The present study relies mainly on prospective EFL teachers because “the field of foreign language education has always been in the forefront of the use of technology to facilitate the language education process” (Lafford and Lafford, 1997). Moreover, language teachers have identified various uses of ICT in the language class, including the use of the web as a tool for obtaining information, project writing, communication among a group of students through e-mail or an online chat-room, web-based course programs and publishing project work (Lee, 2000; Alvine, 2000; Lee, 2003). However, the integration of ICT into instruction has been a major challenge for every teacher because they are expected to be capable of utilizing the extensive capacities of ICT to create more effective teaching and learning activities. If we have any hope of enabling future language teachers to use ICT tools in ways that will capture new learning styles and create pathways for learners living in a digital era, ICT professional development intentions should be met by diagnosing distance learners’ needs in this respect and accordingly ‘re-tooling’ with infrequent curriculum integration.

Therefore, the impetus of this research is to explore the distance prospective EFL teachers’ perceptions of ICT integration by utilizing a questionnaire consisting of questions that are somewhat related to the technology acceptance model (TAM) which was originated by Davis in 1986 (Davis, 1989). The model indicates that perceived usefulness and perceived ease of use are two specific determinants of a user’s acceptance of a technology. Perceived usefulness indicates the user’s perception of the extent that the technology will develop his job performance. This includes decreasing the time for performing the job, more productivity and accuracy. Perceived ease of use refers to the user’s perception of amount of the effort necessary for using the system or the extent to which a user believes that employing a specific technology will be easy. By means of an interview, the study also unearths factors impeding distance prospective EFL teachers’ positive beliefs and their future expectations related to ICT professional development.

**Significance of the study**

This research study is significant for two main reasons. First, it is increasingly important to direct our attention to the centroid shift observed in the field of foreign language teaching through the technology, from the paradigm of CALL-based English language teaching to the new forms of multiple literacies, highlighting the role of ICT (Kellner, 2000; Warschauer, 2000; 2004). Second, teachers’ attitudes towards ICT integration has been a fashionable research topic investigated in many studies in the last decade (Albirini, 2006; Tondeur, van Braak & Valcke, 2007; Aydm, 2007; Teo, Lee, & Chai, 2008). However, distance pre-service EFL teachers’ attitudes towards ICT integration have been neglected in Turkey, where the major educational reforms have been made specific to ICT integration since 1997. Thus, the present study was intended not only to shed more light on distance pre-service EFL teachers’ attitudes toward ICT integration but also to reveal the factors obstructing their positive attitudes. In the following section, distance education in Turkey will be introduced to provide international readers with a basic understanding of the research context.

**Distance education in Turkey**

ICT has been increasingly used in education and has brought up some new disciplines, or study fields, and distance education (DE) is one of the latest forms of education that basically depends on ICT use. Since DE is a modern
approach to delivering instruction, both formal and non-formal educational institutions all around the world and also in Turkey have strikingly been engaged in DE. Although DE can be traced back fifty years in Turkey, it was first mentioned in a meeting in 1927 as a means to increase the number of literate people (Alkan, 1987, p. 91). Unfortunately, it was not applied in Turkey owing to the idea that education is not possible without a teacher and a real classroom atmosphere. It appeared again in the 1970s with a closer meaning of distance education today. In 1982, the Council of Higher Education was established to plan, organise, administer, supervise and regulate all higher education institutions in the country. Belatedly, within the same year, the Open Education Faculty (OEF) was founded in Anadolu University, as it had sufficient infrastructure. Today, the OEF offers a great variety of undergraduate degree programs to more than 750,000 Turkish students across Turkey, Cyprus and Europe (Demiray, 1998).

The Distance English Language Teaching (DELT) undergraduate program was established in 2000 with Basic Education Law numbered 2547 to meet the great need for EFL teachers, especially in primary and secondary schools. This was because English courses were integrated into their curricula in 2000. The DELT undergraduate program lasts for four years and during the first two years, students are provided partially face-to-face instruction for some basic courses such as contextual grammar, oral communication skills, advanced reading and writing, listening and pronunciation and then they have distance education for the last two years. As Bates & Bates (2000) state, DE programs worldwide use a range of technologies that include two-way interactive and one-way non-interactive delivery of courses. Instruction in the distance higher education system in Turkey is based on textbooks, television, CD-ROMs and radio broadcasts, which are, in general, one-way non-interactive delivery methods.

The aforementioned teacher training degree program offers three courses with respect to ICT implementation. Two of them, Computer I and Computer II, aim to improve teacher candidates’ computer competency. The other, Instructional Technology and Material Design, aims to teach them how to use ICT pedagogically in the language teaching and learning processes. Nevertheless, the problem is that these courses are offered at distance through the use of mostly printed materials. Thus, prospective EFL teachers do not have the opportunity to experience ICT which is an inescapable component of powerful language learning and teaching. With this study, we aim to emphasize that prospective EFL teachers at OEF need in-class experience with ICT to be convinced of its usefulness and educational value.

Methodology

Sample characteristics

The study was conducted with the participation of 85 prospective EFL teachers in distance higher education in Turkey. The participants were in the final year of a 4-year program leading to a bachelor’s degree in Teaching English as a Foreign Language. When the participants had distance instruction during the 4th year of their education, they were living in different regions of the country. Each participant was assigned a number during the data analysis procedure because of ethical considerations and they were reminded that this research would not be used for any assessment purposes and that personal details would be kept confidential. The demographic properties of the participants are presented in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Demographic properties of the participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>20-22</td>
</tr>
<tr>
<td>23-24</td>
</tr>
<tr>
<td>25-above</td>
</tr>
<tr>
<td>Computer</td>
</tr>
<tr>
<td>Literate</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>
Instrument and data collection

After an extensive literature review of instruments utilized in different educational backgrounds (Jones and Clarke, 1994; Robertson et al., 1995; Sooknan, 2002; Isleem, 2003; Dudeney and Hockly, 2007), a questionnaire was developed by the researchers to gather data about the perceptions of prospective EFL teachers in distance higher education toward ICT integration, rather than using an existing one. The questionnaire consisted of two parts. The first part asked about personal information such as gender, age and year of education to ensure maximum control of variables (Gay & Airasian, 2000). The second part of the questionnaire contained nine items based on a 5-point Likert scale (from 1=strongly disagree to 5= strongly agree). The development of the instrument was guided by a number of experts working in higher education settings. This panel of experts, including two professors of educational technology, two native experts and two non-native senior EFL teachers, evaluated the instrument for content and face validity and contended that the questionnaire was appropriate and comprehensive for the context of the study. To check the reliability, the instrument was analyzed through the Cronbach’s Alpha Coefficient $\alpha = 0.90$, which showed a high level of reliability.

The questionnaires were administered to the prospective EFL teachers at OEF (N=120) through the Internet because they were in different parts of Turkey during December 2009. The return rate from OEF student teachers was 70.8% (N=85). After rigorous analysis of the collected data, 30 of these prospective teachers were contacted again to get their phone numbers for the interview session. The purpose of the interviews was to cross-check students’ responses to the questionnaire and obtain further information on their responses. 73.3% of the contacted prospective EFL teachers (N=22) agreed to have an interview on the phone. Even though an interview via video chat tools was planned, it was not used since some of the prospective EFL teachers did not have equal opportunities to access any video chat tools. These twenty-two respondents were interviewed throughout three weeks and their responses to the interview questions were transcribed as we had no chance to record their answers. The questions that we posed in the interview were as follows:

1. What obstructs your positive perceptions toward ICT integration?
2. Do you think that learning ICT use at a distance is a disadvantage?
3. What do you suggest to better learn how to integrate ICT?

Data analysis

Data were analyzed using the Statistical Package for Social Sciences (SPSS. 16). The demographic variables for this study were discrete data (nominal and ordinal); therefore, descriptive statistics were utilized to run for frequencies, percentages, mean and standard deviation (Beins, 2004; Heiman, 2001; Sekaran, 2003).

Results

Students' scores on questionnaire

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Standard Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>3.40</td>
<td>0.49</td>
<td>0.05</td>
</tr>
<tr>
<td>Item 2</td>
<td>2.14</td>
<td>0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>Item 3</td>
<td>4.07</td>
<td>0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>Item 4</td>
<td>1.89</td>
<td>0.61</td>
<td>0.06</td>
</tr>
<tr>
<td>Item 5</td>
<td>2.84</td>
<td>0.74</td>
<td>0.08</td>
</tr>
<tr>
<td>Item 6</td>
<td>3.57</td>
<td>0.77</td>
<td>0.08</td>
</tr>
<tr>
<td>Item 7</td>
<td>3.57</td>
<td>0.77</td>
<td>0.08</td>
</tr>
<tr>
<td>Item 8</td>
<td>2.52</td>
<td>0.50</td>
<td>0.05</td>
</tr>
<tr>
<td>Item 9</td>
<td>2.52</td>
<td>0.50</td>
<td>0.05</td>
</tr>
<tr>
<td>Overall attitude</td>
<td>2.77</td>
<td>0.19</td>
<td>0.02</td>
</tr>
</tbody>
</table>
As seen in table 2, the results of the questionnaire indicated that although prospective EFL teachers in distance higher education viewed ICT as a tool to help them to learn many new things (scoring 4.07 on item 3 in the questionnaire), most of these prospective EFL teachers (a) disliked using ICT (item 1, M=3.40; SD=0.49), (b) viewed using ICT in class as time-wasting (item 2, M=2.14; SD=0.35), intimidating or threatening them (item 4, M=1.89; SD=0.61), (c) felt uncomfortable when working with ICT in class (item 5, M=2.84; SD=0.74), (d) were unwilling to learn more about employing ICT in class (item 6, M=3.57; SD=0.77), (e) disregarded the power of ICT as a device to develop their teaching practice (item 7, M=3.57; SD=0.77), (f) considered modifying the curriculum to integrate ICT as a difficult task (item 8, M=2.52; SD=0.50) and (g) stated that ICT would break down too often to be of very much use (item 9, M=2.52; SD=0.50). That is, they generally displayed negative attitudes towards ICT integration by scoring lower than 4 on most of the items on the five point scale.

Interviews

Factors obstructing positive perceptions toward ICT integration

In response to the first question regarding the factors impeding their positive perceptions of ICT use in the language learning or teaching process, the participants expressed mainly three obstacles which were (1) the lack of exposure to lessons fully designed with ICT tools, (2) an exam-driven system and (3) studying to learn only what is to be tested. The following selected quotations (presented verbatim) are representative of the views expressed:

“I have never seen a lesson fully designed with ICT tools, so how can I conduct an ICT integrated lesson for my students?” (Participant 21)

“In an exam-based system, it is very difficult to learn something practical. I really wish to teach my students an Internet-based lesson but how?” (Participant 3)

“I learn what I am required to learn. If I am required to learn about ICT, I will study it as it is tested in the exams.” (Participant 9)

Learning ICT use at a distance: a disadvantage?

The participants were asked whether or not they thought that learning ICT use at a distance was a disadvantage. The majority of them stated that learning ICT use at a distance was a disadvantage for them due to the lack of opportunities to try ICT use, the need to do practice in a technology laboratory, and the lack of educational technology teachers. The following quotations illustrate prospective EFL teachers’ views on learning ICT use at a distance:

“We are not as lucky as our peers in formal education because we have no opportunities to try ICT use.” (Participant 9)

“Learning ICT use at a distance is not easy for us. We need to practice in a technology laboratory in a face-to-face classroom setting to better understand how to use ICT in a language classroom.” (Participant 17)

“Learning ICT use at a distance is problem-causing for us because we have no educational technology teachers who can show us practically how to make use of ICT in a real classroom environment” (Participant 13)

Suggestions to better learn how to integrate ICT

In response to the third question related to suggestions to better learn how to integrate ICT, participants stressed the prominence of (a) training specific to methodologies and practices of ICT integration, (b) ICT-integrated sample lessons conducted in face-to-face classroom environments by educational technology specialists and (c) providing computer laboratories in central cities of our country to increase prospective EFL teachers’ self-efficacy and decrease their anxiety about using ICT. From the interviews with the participating prospective EFL teachers, some of the suggestions made are:
“Raising our computer skills in Computer I and Computer II lessons does not guarantee our using ICT because we need to have clearer ideas of how to apply these ICT skills in teaching through training in methodologies and practices of ICT integration.” (Participant 7)

“We can learn how to use ICT in the language classroom by being exposed to ICT integrated sample lessons conducted in face-to-face classroom environments by educational technology specialists.” (Participant 12)

“Even if we are distance learners, we should be provided with computer laboratories in, at least, central cities of our country so that we can increase self-efficacy and decrease our anxiety about using ICT skills.” (Participant 9)

In this research study, the results of the quantitative data were closely connected with the results of the qualitative data. Firstly, based on the results of the questionnaire, it was indicated that most of the prospective EFL teachers in distance higher education generally exhibited negative attitudes towards ICT integration. Related to these results, the prospective EFL teachers’ responses to the first and the second questions of the interview unearthed that lack of exposure to lessons fully-designed with ICT tools, lack of opportunities to try ICT use, the need to practice in a technology laboratory, lack of educational technology teachers, exam-driven education system and studying to learn only what is to be tested were some of the underlying reasons for the prospective EFL teachers’ negative perceptions of ICT use in the language learning or teaching process.

Secondly, based on the results of the questionnaire, it was also indicated that prospective EFL teachers in distance higher education viewed ICT as a tool to help them to learn many new things. In parallel to this result, the prospective EFL teachers’ responses to the third question of the interview revealed that they gave importance to (a) training with respect to methodologies and practices of ICT infusion, (b) ICT-infused model lessons offered in face-to-face classroom environments by educational technology specialists and (c) establishing computer laboratories in central cities of our country to maximize prospective EFL teachers’ self-efficacy and decrease their anxiety about using ICT.

Discussion

Researchers have long argued for the inclusion of ICT in foreign language teacher training (e.g., Blake, 2001; Schrier, 2001; Lord and Lomicka, 2004). In this respect, some studies examined the actual content of technology training modules (e.g., Johnson & Johnson, 1999; Rava and Rossbacher, 1999; Hargrave and Hsu, 2000), while others explored teachers’ attitudes toward and perceptions of technology and its pedagogical uses (e.g., Albirini, 2006; Tondeur, van Braak & Valcke, 2007; Aydin, 2007; Teo, Lee, & Chai, 2008). Unlike the studies done in formal higher education settings that documented positive teacher attitudes toward the use of ICT as teaching tools (e.g., Le and Le, 1999; Brandl, 2002), this research study done in a distance higher education context unearthed negative teacher attitudes toward the use of ICT in language teaching. Lack of exposure to lessons fully-designed with ICT tools, lack of opportunities to try ICT, the need to practice in a technology laboratory, lack of educational technology teachers, an exam-driven education system and studying to learn only what is to be tested were some of the underlying reasons for the prospective EFL teachers’ negative perceptions of ICT use in the language learning or teaching process. These findings provide evidence in support of Schoepp (2005), Sicilia (2005), Toprakçı (2006), Albirini (2006), Balanskat et al. (2006), Özden (2007) who indicate that lack of effective training is the most frequently encountered barrier to successful ICT integration.

The results of this study also apparently show that despite having basic computer skills, prospective EFL teachers in distance education were not confident in using the technology to improve their own productivity and bring about a pedagogical change in their teaching methods. Hence, the need for conducting undergraduate courses on the basis of ‘learning by doing’ is evident. At this juncture, distance teacher training programs should reconsider their training approaches in order to provide prospective EFL teachers in distance higher education with more conducive and non-threatening learning environments so that these prospective EFL teachers can experience ICT tools, gain confidence and increase self-efficacy for technology integration (Wang et al., 2004; Teo, 2008).

Moreover, this research study reveals that prospective EFL teachers in distance higher education, despite having negative attitudes toward ICT integration and not utilizing it in the classroom, viewed ICT as a tool to help them to learn many things. This finding is in line with the finding of Smeets (2005) who stressed that most teachers do not
utilize the potential of ICT to maximize the quality of learning environments, although they value this potential quite significantly. At this point, it should be emphasized that prospective teachers’ beliefs are one step toward developing effective professional development (Mueller et al., 2008) since their beliefs about the role of ICT for learning and teaching are important in teachers’ pedagogical reasoning and they influence teachers’ behaviour (Webb and Cox, 2004) and their beliefs often limit their efforts to integrate ICT into classroom practices (Pelgrum, 2001).

Finally, it is obvious that unless teachers perceive the new technologies as valuable, they will be unwilling or unable to use them meaningfully. Accordingly, the importance of professional development for successful infusion of ICT in the classroom is strongly endorsed in the literature (e.g., O’Brien et al., 1999). Baylor and Ritchie (2002) found that support for professional development and the level of technology are embraced in the classroom. As a last remark, given participating prospective teachers’ limited experience in employing educational ICT and their lack of exposure to ICT-infused teaching methods, the points they made seem commonplace when compared with those from other findings (e.g., Alvine, 2000; Pope and Golub, 2000).

Implications, limitations and conclusion

The educational implications of the study

Distance EFL teacher training programs should reevaluate their teaching methods and give importance to training specific to methodologies and practices of ICT integration, ICT-integrated sample lessons conducted in face-to-face classroom environments by educational technology specialists and providing computer laboratories in central cities of our country to increase prospective EFL teachers’ self-efficacy and decrease their anxiety about using ICT skills.

Limitations and future directions

This study has three main limitations which are as follows:

1. This study contains prospective EFL teachers who are in distance English teacher training departments. Because of this reason, the results of this study cannot be generalized to include prospective EFL teachers in formal education contexts or prospective teachers in other departments of OEF.

2. There are approximately 7,500 prospective EFL teachers in the DELT program. However, we were able to reach only 85 of these prospective EFL teachers. Hence, the results that we obtained cannot be generalized to the whole population.

3. The data gathered through interviews is limited to the respondents who voluntarily participated in these interviews. It is likely that interviewing with more respondents could help us gain more insights related to this subject.

Accordingly, more qualitative studies should be directed to explore how prospective language teachers perceive ICT integration in language instruction. Furthermore, studies that compare and contrast perceptions of prospective EFL teachers in formal education contexts with those of prospective EFL teachers in distance education can be done.

Conclusion

ICT presents a powerful learning environment for learners in the classroom. Many countries make investments in ICT integration as ICT is viewed as an effective tool for renewing educational practice in any field. Because teachers are the main characters to employ ICT in educational contexts, they should be trained in how ICT can be integrated into the teaching process. In the present study, almost all of the respondents expressed their preference for face-to-face training instead of distance training of ICT use in the language teaching process and most future language teachers are very traditional in that they are less enthusiastic about learning ICT skills.

Unfortunately, many higher education institutions cannot provide pre-service teachers with positive technological experiences. Thus, pre-service teachers with negative ICT perceptions cannot transfer their ICT skills to their students and stimulate them to deploy ICT when they start teaching. However, if we aim to make future language
teachers deploy ICT tools in language instruction, we should investigate pre-service teachers’ perceptions of ICT integration since perceptions are good indicators of intentions and future behaviors in future language teaching and diagnose pre-service EFL teachers’ needs in relation to the integration of ICT into instruction.

As teacher training is at the first zenith and as ‘the heat is on’ for those who prepare educators for effective use of ICT (Knierzinger et al., 2002), for more effective solutions, special attention should be paid to revising and updating the curricula, equipment and educational materials on a permanent basis in the faculties of education. As to distance higher education in Turkey, it is to be noted that ICT can create a better teaching and learning environment in schools (Akkoyunlu and Orhan, 2001) as long as prospective teachers are trained well through a curriculum rescued from traditional behaviorist approach domination and the curricula, equipment and educational materials are designed with a more internationally accepted and fast-growing educational model based on practicing and experiencing.

Unfortunately, achieving this aim in a distance education setting is rather difficult as the students live in separate places. Considering the time and place independency feature of distance education, we suggest two major ideas:

1. The course related to ICT integration, Instructional Technology and Material Design, should be offered within the first two years of the program, during which prospective teachers can come to central cities to join the partially face-to-face sessions. Thus, prospective EFL teachers may decide on what facets of ICT they need to explore and learn them by doing in the classroom.
2. If the above-mentioned course is offered in the last year of the program (as is the case in the current system), computer laboratories should be established in big cities, if possible in all the cities of the country, so that distance learners can attend face-to-face sessions at least once a month. To supplement this continuous training, prospective EFL teachers can be required to prepare ICT-integrated lesson plans and then present them in their practice schools, whereby they can get feedback from their supervisors.

In doing so, certain problems should be solved such as lack of computer laboratories for distance learners, insufficient staff in the area of ICT training, insufficient access to ICT tools and lack of research in this field. As Altun (2007) suggests, there is an urgent need for relevant empirical research in line with the integration process in Turkey.

References


Toprakci, E. (2006). Obstacles at integration of schools into information and communication technologies by taking into consideration the opinions of the teachers and principals of primary and secondary schools in Turkey. *Journal of Instructional Science and Technology (e-JIST), 9*(1), 1-16.


APPENDIX A. Questionnaire for Distant Students of English Language Teaching Department in Turkey

With this survey, we hope to collect data about the perceptions of prospective EFL teachers at distance higher education context toward ICT integration. The questionnaire consists of two parts. In the first part, you are asked to fill in your personal details. In the second part, you are requested to read nine statements below and then rank the items between 1 (strongly disagree) to 5 (strongly agree).

(1=totally disagree  2= disagree  3= no strong opinion  4= agree  5= strongly agree)

Part I
Age: ______       Gender: _______       Year of education ________
Can you use computers? ___________________________________________________
Have you taken any course related to ICT use? _________________________________
If yes, please specify these courses: __________________________________________

Part II

<table>
<thead>
<tr>
<th>Attitudes to Information and Communication Technology</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I enjoy using Information and Communication Technology.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I think using ICT saves time in class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I know that ICT can help me to learn many new things.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Using ICT does not intimidate or threaten me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I feel very confident when it comes to working with technology in class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I want to learn more about using ICT in class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I believe that ICT can really improve my teaching practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Changing the curriculum to integrate ICT is possible.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. ICT does not break down too often to be of very much use.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thank you very much for your collaboration.

Murat Hismanoglu
The Impact of Recurrent On-line Synchronous Scientific Argumentation on Students’ Argumentation and Conceptual Change

Chien-Hsien Chen and Hsiao-Ching She

Institute of Education, National Chiao-Tung University, Taiwan // ptm1122@gmail.com // hcshe@mail.nctu.edu.tw

ABSTRACT

This study reports the impact of Recurrent On-Line Synchronous Scientific Argumentation learning on 8th grade students’ scientific argumentation ability and conceptual change involving physical science. The control group (N=76) were recruited to receive conventional instruction whereas the experimental group (N=74) received the Recurrent On-Line Synchronous Scientific Argumentation program for about 25 physical science class periods of 45 minutes each, which is about one third of the physical science class periods in a semester. Results indicate that the experimental group significantly outperformed the conventional group on the post-Physical Science Conception Test and the Physical Science Dependent Argumentation Test. The quantity and quality of scientific arguments that the experimental group’s students generated, in a series of pre- and post-argumentation questions, all improved across the seven topics. In addition, the experimental group’s students successfully constructed more correct conceptions from pre- to post-argumentation questions across the seven topics. This clearly demonstrates that the experimental group’s students’ argumentation ability and conceptual change were both facilitated through receiving the Recurrent On-Line Synchronous Scientific Argumentation program.

Keywords

Scientific Argumentation, Conceptual Change, On-line Synchronous argumentation, Physical science, Recurrent on-line learning, 8th grade students

Introduction

The need to educate our students and citizens about how we know and why we believe in the scientific worldview has become increasingly important. It is no longer sufficient to merely deal with what we know (Driver et al., 1996; Millar & Osborne, 1998). Osborne et al. (2004) further pointed out that such a shift requires a new focus on how evidence is used in science for the construction of explanations, that is, on the arguments that form the links between data and the theories that science has constructed. More specifically, the construction of arguments is a core discursive activity of science (Osborne et al., 2004). Scientific discursive practices such as assessing alternatives, weighing evidence, interpreting texts, and evaluating the potential validity of scientific claims are all seen as essential components in constructing scientific arguments, which also are fundamental in the progress of scientific knowledge (Latour, 1987). In short, argumentation is a collective cognitive development process which involves using evidence to support or refute a particular claim, coordinating the claims with evidence to make an argument, forming a judgment of scientific knowledge claims, and identifying reliable and consensual scientific knowledge. Several studies show that educational support of argumentation may foster students’ argumentation ability (Jiménez-Aleixandre, & Rodriguez, 2000; Kuhn et al., 1997) and improve scientific knowledge (Zohar & Nemet, 2002).

Most of the argumentation studies were conducted in the classroom for a very short period of time and were not able to improve students’ argumentation efficiently. The authors feel that it is necessary to provide students with the opportunity to argue effectively with recurrent opportunities and for a longer period of time in order to improve the quality of their argumentation. Osborne et al. (2004) suggested that developing argumentation in a scientific context is far more difficult than enabling argumentation in a socio-scientific context. Students generally considered physical science to be difficult to learn. Though it is rather difficult to improve argumentation in a science context, we believe it is important to provide students with the recurrent opportunity to learn and use argumentation in the context of physical science.

The constructivist view of learning highlights the significance of the individual learner’s prior knowledge in subsequent learning (Driver & Bell, 1986). Cobern (1993) shares the similar idea of learning as a process wherein an individual is actively involved in linking new ideas with current ideas and experience. Learning by construction and involving changes is similar to the idea that the construction of new knowledge takes place at a construction site consisting of existing structures built on a foundation (Cobern, 1993). The notion of conceptual change involves the restructuring of relationships among existing concepts and often requires the acquisition of entirely new concepts. The students who learn something are the ones who understand a new idea, judge its truth value, judge its
consistency with other ideas, and are willing to change their minds to accept it. It is recognized that learning does not take place in a social vacuum (Driver, 1995). Driver (1995) indicated that whether or not an individual’s ideas are affirmed and shared by others in classroom exchanges affects how the knowledge construction process is shaped. The nature of argumentation has the potential to contribute to the collective development and judgment of scientific knowledge claims and the identification of reliable and consensual descriptions of nature (Kolsto & Ratcliffe, 2008, p.117). Much effort and many studies have focused on fostering students’ conceptual change from the constructivist viewpoint (She, 2004; She & Lee, 2008; She & Liao, 2010; Hewson & Hewson, 1988; Venville & Treagust, 1998). However, none of them have tried to include argumentation in fostering conceptual change.

Obviously, argumentation has great potential for fostering students’ communication skills in order to interchange perspectives and meanings. Assessing alternatives, weighing evidence, interpreting texts, and evaluating the potential validity of scientific claims are all seen as essential components in constructing scientific arguments. But how can argumentation successfully foster students’ conceptual change? According to the previous conceptual change studies, the following major characteristics are important for successful conceptual change: (1) Creating dissonance, which can raise students’ awareness about their own conceptions and provide an opportunity for them to experience the dissonance and become further dissatisfied with their own conceptions (She, 2004; Posner et al., 1982). (2) Challenging students’ beliefs about science conceptions (She, 2004; Vosniadou & Brewer, 1987). (3) Providing plausible mental structures for students to reconstruct more scientific conceptions (She, 2004). (4) Actively engaging students in the process of conceptual change (Hewson & Hewson, 1983; She, 2004). (5) Actively involving students in group discussions to shape their knowledge construction process and changing conceptions (Driver, 1995; Venville & Treagust, 1998). Therefore, the ideas of successful conceptual change described above were taken into consideration during our design of argumentation activity in order to optimize scientific learning.

Though there is a substantial body of research showing that instructors tend to adopt their conventional instruction into online courses, however, Scagnoli, et al. (2009) suggested that simply changing face-to-face courses to an online environment can’t confirm the same success. Additionally, consensus has not been reached regarding whether online learning is more effective than conventional instruction on students’ academic achievement. Many studies suggest no difference in academic achievement scores following on-line learning and conventional courses (Delfino & Persico, 2007; Russell, et al. 2009). Larson and Sung (2009) further demonstrated that there are no significant differences on exam scores when comparing online, blended and face-to-face instruction. On the other hand, the majority of articles reported that online learning is better than traditional learning with the focus on the perspectives of engagement or social situations, pedagogical characteristics, and satisfaction (Larson & Sung, 2009; Menchaca, 2008; Wuensch et al., 2009) instead of focusing on students’ learning outcome. One study demonstrated that, for a wellness course, the online learning group’s levels of achievement were significantly higher than those of the traditional face-to-face learning group (Lim et al., 2008). Kirtman (2009) reported a contrary result that students who received traditional instruction performed significantly better than the students who received online instruction, in both mid-term and final exams. Salcedo (2010) further demonstrated that students who received traditional instruction for foreign language classes performed better on three out of four assessments than did the students who received online instruction. As the consensus still remains unclear, we are interested in exploring whether or not students receiving the on-line scientific argumentation course perform better than the conventional group. From the point of view of science educators, we claim that it is very difficult to bring about conceptual change and argumentation unless the instructional design is based on well-developed conceptual change and argumentation theories and models (Yeh & She, 2010). Though a few studies have proposed their on-line argumentative learning environment for promoting students conceptual development and conceptual change (Ravenscroft, 2000, 2007), they lack empirical evidences to prove their effectiveness. Thus, this study attempts to explore whether or not students who received the On-Line Synchronous Scientific Argumentation learning would outperform a conventionally educated group of students in their conceptual change and scientific argumentation.

Sandoval and Reiser (2004) suggest that online learning environments can provide excellent support for students constructing their scientific explanations and knowledge negotiation process in argumentative writing. Synchronous communication can deliver a higher degree of elaboration and construction of arguments as students work on a common shared artifact (De Vries et al., 2002; Janssen et al., 2006). Our study specifically designed a synchronous argumentation Web-based learning environment to provide students with the opportunity to argue with their group in real time and to create a higher degree of elaboration and construction of arguments.
Constructing a good argument is not a simple task and we believe that guidance and support would help students to scaffold and build their sense of an effective argument. On-line argumentation provides the advantage of allowing students to see arguments and counterarguments on the screen, which supports them in refining their argumentation (Kirschner et al., 2003). Wray and Lewis (1997) have indicated that the use of “writing frames” would support the process of writing and provide vital clues as to what is needed. Osborne et al. (2004) indicated that stems provide students with prompts to construct their argument in a coherent manner and within a writing frame, which then can be used as a structure for producing a written argument. Therefore, we specifically programmed our learning environment to provide students with writing frames of five argumentation components to scaffold their arguments in science learning.

The current study specifically designed On-Line Synchronous Scientific Argumentation learning to provide a recurrent opportunity for middle school students to engage in argumentation for about one third of the physical science class periods in a semester. The strategy is to provide students with writing frames of five argumentation components to scaffold their arguments in the On-Line Synchronous Scientific Argumentation learning. We believe that it is a promising direction, taking consideration of conceptual change aspects into the design of a series of pre-, experiment-related, and post-argumentation activities.

**Research Questions**

Three major research questions were examined in the study in order to measure the effectiveness of On-Line Synchronous Scientific Argumentation learning. The first question explored whether On-Line Synchronous Scientific Argumentation learning was more effective than conventional instruction in facilitating students’ conceptual change as well as scientific argumentation in physical science. Second, examine the quantity and quality of scientific arguments that experimental group’s students generated in a series of pre- and post-argumentation questions across a semester. Third, explore the nature and extent of conceptual change from pre- to post-argumentation question that the experimental group’s students made across a semester. In addition, the relationship between scientific conceptual change and argumentation ability was examined.

**Designs and Characteristics of the Recurrent On-Line Synchronous Scientific Argumentation learning**

Recurrent On-Line Synchronous Scientific Argumentation learning is designed to provide recurrent argumentation opportunities for students learning physical science, replacing the regular physical science in middle school. Therefore, the five units of seven topics were chosen from the current middle school physical science mandatory content and standards. The current study reported the effects of implementing seven topics for physical science: chemical reaction (1 and 2), acid and base (1 and 2), oxidation and reduction, organic substances, and friction. Seven topics of physical science were used in this study. Each unit generally covers two or three main topics, for instance unit 1 on chemical reaction covers the influence of the contacting area on the rate of chemical reaction, and the influence of concentration on the rate of chemical reaction. Each topic is specifically designed a pre-argumentation question and an experiment-related argumentation question was focused on the core concepts of the pre-argumentation question that they argued (figure 1). Students were asked to provide reasons for the argument and went to an actual laboratory to carry out their experiments based upon their hypothesis and experimental design. The same post-argumentation question was given for students to argue again after finishing the laboratory work.

**Facilitate students’ conceptual change**

To facilitate students’ conceptual change, each topic is specifically designed to initiate a pre-argumentation question, followed by an experiment-related argumentation question, the activity of carrying out the experiment in the laboratory, and finally a post-argumentation question. Students would be exposed to different ideas which may be different from their own during the pre-argumentation and experiment-related argumentation question. After they carry out the experiment and receive the result from the experiment, dissonance is created and they build a plausible mental structure if the result is different from their prediction. The same post-argumentation question is given for students to argue again after finishing the laboratory work. Post-argumentation provides them an opportunity to reconstruct their mental structure according to the experiments they have visualized, arguing with peers, exchanging
conceptions, justifying their belief, and further modifying their original conceptions throughout the process. This process is intended to encourage students to reconstruct their scientific conceptions through a series of pre-argumentation, question-testing argumentation questions, laboratory activities, and post-argumentation.

Facilitate students’ argumentation ability

In order to promote students’ argumentation ability, the Recurrent On-Line Synchronous Scientific Argumentation learning environment has tools specifically designed for students to use while they are participating in argumentation. In order to facilitate students’ ability to produce a good written argument, our interface specifically designed two layers of templates for them to use. The first layer provides the definition and choices of five components of argumentation: data, claim, warrant, backing, and rebuttal; the second layer provides three or four writing frames for each component of argumentation (figure 2).

Figure 1. On-line pre-argumentation discourse and first layer of template
The writing frame is intended to provide guidance and support that will help students in constructing a good argument. The following stems were provided: “I think/believe..., because...; The reason why I agree with...argument, is because the evidence of......; I do not agree with...my reason is...” Students need to choose one of the components of argumentation first and then choose one from three or four writing frames that they feel appropriate to share their argument. The learning environment provides the advantage of real time argumentation, so students can receive prompt rebuttals to their arguments which can better retain their interest and thus make learning more effective.

Method

Participants and procedures

A total of 150 eighth grade students, recruited from four classes of a middle school, participated in this study. Two classes of students (74) received the Recurrent On-Line Synchronous Scientific Argumentation learning (experimental group) and the other two classes of students (76) received conventional instruction (control group). The experimental group’s students were further divided into 12 groups, with an average of six students assigned to
each group. The experimental group received the seven topics of the physical science Recurrent On-Line Synchronous Scientific Argumentation learning for one semester, with 25 class periods, each class lasting about 45 minutes. This was about one third of the physical science class periods over a semester. The teacher introduced the five components of argumentation to the students and used it in the classroom for one class before the students received the Recurrent On-Line Synchronous Scientific Argumentation learning and assessment. The conventional group students went through the same content of physical science in traditional instruction and traditional laboratory work.

All students were administered the two-tier Physical Science Conception test (PSCT) and the Physical Science Dependent Argumentation Test (PSDAT) before and one week after learning. In addition, the experimental group students’ on-line scientific argumentation process also was collected to determine the quality and quantity of students’ argumentation and scientific conceptions they held before and after learning from the OLSA.

**Instruments**

*Physical Science Conception test (PSCT)*

The PSCT is a two-tier multiple choice diagnostic instrument that was developed to measure the degree of students’ conceptual change in physical science conceptions. The content validity was established by the same panel of six evaluators, ensuring that the items were properly constructed and relevant to the seven topics of physical science Web-learning materials that we developed. There are five items for each topic, and each item contains two tiers. In the first tier, students are required to choose the correct scientific concepts, while in the second tier they choose the correct reason for choosing these specific concepts. There are 35 items and each item has two tiers. Students need to answer both tiers of each question correctly in order to receive one point, so the highest possible score is 35. The Cronbach $\alpha$ of ADRT was 0.86 for the pre-test and 0.92 for the post-test.

*Physical Science Dependent Argumentation Test (PSDAT)*

The PSDAT is a two-tier multiple choice diagnostic instrument that was developed to measure the degree of students’ argumentation ability involving physical science conceptions. There are five scenarios, covering five units of seven topics. Each scenario includes the contextual background and argumentation discourses. There are five questions under each scenario, for a total of 25 questions. Each question contains two tiers. The first tier of each question requires the student to identify a specific statement from the argumentation discourses at scenario as a correct data, claim, warrant, backing, or rebuttal, respectively, and justify why they chose that specific statement as a correct data, claim, warrant, backing or rebuttal. The content validity was established by the same panel of six evaluators, ensuring that the items were properly constructed and relevant to the five units of the OLSA physical science learning program. There are 25 items covering five units. Students need to answer both tiers correctly in order to receive one point, so the highest possible score is 25. The Cronbach $\alpha$ of PSDAT was 0.91 for the pre-test and 0.92 for the post-test.

*Qualitative Analysis of On-line scientific argumentation*

The qualitative data collected from students’ on-line scientific argumentation was analyzed from two perspectives. Each statement generated by an individual was classified into two different levels of claim, warrant, backing and rebuttal, respectively. Data is considered to be non-argumentative statements. A level 1 claim is an argument consisting of a claim without any data or fact. A level 2 claim is an argument consisting of a claim with data or fact. A level 1 warrant is an argument consisting of a theory or principle without connection to the claim, or one which does not clearly describe the theory. A level 2 warrant is an argument consisting of a claim with a clearly described theory or principle. A level 1 backing is an argument only consisting of a backing without any connection to claim/warrant, or one which does not clearly describe the connection among them. A level 2 backing is an argument consisting of a claim with backing, and or with data or warrant. A level 1 rebuttal is an argument consisting of a weak rebuttal without clear explanation. A level 2 rebuttal is an argument consisting of a claim with a clearly identifiable rebuttal (Table 1). The cross-coder reliability is 0.91.
In addition, students’ on-line scientific argumentation discourses were analyzed from a conceptual change perspective. Each argumentation statement was determined to be correct, partially correct, or incorrect; and the comparison between the correctness of pre-argumentation and post-argumentation is further performed by the t test. The quality of conceptual change also was presented to see how students change their conceptions from pre- to post-argumentation questions across seven topics. The cross coder reliability is 0.95.

Table 1. Analytical Framework used for determining the quality of argumentation

<table>
<thead>
<tr>
<th>Components</th>
<th>Levels</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim</td>
<td>Level 1</td>
<td>An argument only consists with a claim without any data or fact.</td>
<td>The greater the concentration, the faster the reaction is.</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>An argument consists of a claim with data or fact.</td>
<td>I saw that the greater the concentration of HCl, the faster the reaction with marble is. Thus I think that the greater the concentration, the faster the reaction is.</td>
</tr>
<tr>
<td>Warrant</td>
<td>Level 1</td>
<td>An argument only consists with a theory or principle without connection to the claim, or not clearly describes the theory.</td>
<td>The more molecules there are, the greater the opportunity for collision.</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>An argument consists of a claim with theory or principle.</td>
<td>The greater the concentration is, the faster the reaction is. It is because the more molecules there are, the greater the opportunity for collision.</td>
</tr>
<tr>
<td>Backing</td>
<td>Level 1</td>
<td>An argument only consists with a backing without any connection to claim/warrant, or not clearly describe the connection among them.</td>
<td>I agree with David’s idea, because I had a similar experience that producing oxygen experiment with high concentration of hydrogen peroxide.</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>An argument consists of a claim with backing, and or with data or warrant.</td>
<td>I support Ann’s idea, because I have done the concentration experiment (HCl react with marble), which proves that the greater the concentration, the faster the reaction is. So there is greater intensity of the molecular collisions.</td>
</tr>
<tr>
<td>Rebuttal</td>
<td>Level 1</td>
<td>An argument only consists of a weak rebuttal and without clearly explanation.</td>
<td>I do not agree with Thomas’s idea, because that some person who drink high concentration wine would not get drunk at all.</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>An argument consists of a claim with a clearly identifiable rebuttal.</td>
<td>I disagree with Jim’s idea that the lower the concentration is, the faster the reaction is. The lower the concentration, the smaller the amount of molecules, thus the lower the opportunity for collision.</td>
</tr>
</tbody>
</table>

Results

ANCOVA analysis of the Physical Science Conception test (PSCT)

The two-tier PSCT was developed to measure the degree of students’ conceptual change in physical science conceptions. One-factor ANCOVA was conducted to examine the effects of instructional approaches using post-PSCT scores as the dependent measures, and students’ pre-PSCT scores as the covariate. The results of the one-factor ANCOVA: specifically, instructional approaches (F=4.86, p= 0.029) reach a statistically significant effect on the performance of post-PSCT. In summary, the OLSA group outperformed the traditional group on post-performance of Physical Science Conception test.

Multivariate analysis of the Physical Science Dependent Argumentation Test (PSDAT)

One-factor ANCOVA was conducted to examine the effects of instructional approaches using post-PSDAT scores as the dependent measures, and students’ pre-PSDAT scores as the covariate. The results of the one-factor ANCOVA:
specifically, instructional approaches (F=7.28, p= 0.008) reach a statistically significant effect on the performance of post- and retention-PSDAT. In summary, the OLSA learning group outperformed the traditional group on post-performance of Physical Science Dependent Argumentation Test.

Multiple regression analysis

This section examines the relationship between students’ degree of conceptual change and their scientific argumentation ability. Therefore, the stepwise regression method was used to explore whether the pre- PSDAT or pre-PSCT test would be most important for predicting the post-PSDAT scores. Results indicated that the best single predictor for post-PSDAT scores was the pre-PSCT, followed by pre-PSDAT scores. The standardized regression coefficient for pre-PSCT, and pre-PSDAT were 0.41 and 0.31. Together pre-PSDAT and pre-PSCT accounted for 38.0% of the variance in post-PSDAT scores.

The Quantity and Quality of On-Line Scientific Argumentation

The experimental group’s student on-line scientific argumentation learning process was analyzed in two aspects: nature and extent of argumentation ability and of conceptual change. The quality and quantity of students’ argumentation and conceptual change were presented in the following in order to manifest the nature and extent of experimental group’s on-line scientific argumentation process.

Argumentation ability

All argumentation questions were designed to require 10-15 minutes for students to argue. With an average of six students in a group, the mean frequency of arguments generated by each group in each question increased progressively from 7.38 to 18.77 arguments during the 10-15 minutes from topic 1 to 7 (Figure 3).

![Figure 3. Distribution of mean frequency of arguments generated by each groups’ students across seven units](image)

Repeated measures of ANOVA were used to examine any increases in mean frequency of arguments from topic 1 to topic 7. The mean frequency of arguments generated by each student in each question significantly increased from
1.17 to 3.04 from topic 1 to 7 (F=30.74, p<0.0001) (Table 2). Clearly, the group argumentation and individual students’ argumentation pattern are similar. The post-hoc comparisons indicated that the number of arguments generated by each student is statistically significantly greater when comparing later topics with earlier topics. This clearly demonstrates that students’ ability to generate arguments indeed increased from topic 1 to topic 7 across the semester.

Table 2. Repeated measures of ANOVA of arguments generated by each student across seven topics

<table>
<thead>
<tr>
<th>Topic</th>
<th>M</th>
<th>SD</th>
<th>ANOVA</th>
<th>Post hoc comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>F value of repeated measures</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td>1.17</td>
<td>.61</td>
<td>30.74***</td>
<td>(p=.000)</td>
</tr>
<tr>
<td>Topic 2</td>
<td>2.20</td>
<td>2.20</td>
<td>2&gt;1***</td>
<td></td>
</tr>
<tr>
<td>Topic 3</td>
<td>2.68</td>
<td>2.05</td>
<td>3&gt;1***, 3&gt;4**, 3&gt;5**</td>
<td></td>
</tr>
<tr>
<td>Topic 4</td>
<td>2.05</td>
<td>1.16</td>
<td>4&gt;1***</td>
<td></td>
</tr>
<tr>
<td>Topic 5</td>
<td>1.91</td>
<td>1.05</td>
<td>5&gt;1***</td>
<td></td>
</tr>
<tr>
<td>Topic 6</td>
<td>2.38</td>
<td>1.22</td>
<td>6&gt;1***, 6&gt;4*, 6&gt;5***</td>
<td></td>
</tr>
<tr>
<td>Topic 7</td>
<td>3.04</td>
<td>1.38</td>
<td>7&gt;1***, 7&gt;2**, 7&gt;4***, 7&gt;5***</td>
<td></td>
</tr>
</tbody>
</table>

* p<.05, ** p<.01, *** p<.001; N=74.

Each statement was categorized into two levels of claim, warrant, backing and rebuttal arguments. With an average of six students in a group, the mean frequency of claim, warrant, backing and rebuttal arguments generated by each group in each question from topic 1 to 7 ranged 2.29-13.3, 0.63-4.58, 0.75-4.27, and 0.18-1.77 (Figure 4).

Figure 4. Distribution of mean frequency of claim, warrant, backing and rebuttal arguments generated by each groups’ students across seven units.
Table 3 shows that the mean frequency of claim, warrant, backing and rebuttal arguments generated by each student for each question from topic 1 to 7 ranged from 0.37-2.16, 0.10-0.97, 0.12-0.69, and 0.03-0.29. The increase of arguments was found to be statistically significant when comparing earlier topics with later topics through the use of repeated measure of ANOVA in all aspects, regardless of claim, warrant, backing and rebuttal \( F(\text{claim})=32.69, p <0.0001; F(\text{warrant})=29.95, p <0.0001; F(\text{backing})=11.63, p <0.0001; F(\text{rebuttal})=6.06, p <0.0001 \). The post-hoc comparisons indicated that the frequency of arguments is statistically significantly greater when comparing later topics with earlier topics, regardless of claim, warrant, backing, and rebuttal in general.

Table 3. Multivariate Analysis of Covariance (MANCOVA) of Claim, Warrant, Backing, and Rebuttal Arguments Generated by Each Student across seven topics

<table>
<thead>
<tr>
<th>Units</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Total</th>
<th>F*</th>
<th>Post hoc comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAIM</td>
<td></td>
<td></td>
<td></td>
<td>32.69***</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td>.31</td>
<td>.33</td>
<td>.05</td>
<td>.13</td>
<td>.37</td>
</tr>
<tr>
<td>Topic 2</td>
<td>.47</td>
<td>.74</td>
<td>.08</td>
<td>.19</td>
<td>.55</td>
</tr>
<tr>
<td>Topic 3</td>
<td>1.81</td>
<td>1.40</td>
<td>.35</td>
<td>.81</td>
<td>2.16</td>
</tr>
<tr>
<td>Topic 4</td>
<td>1.07</td>
<td>.85</td>
<td>.33</td>
<td>.48</td>
<td>1.41</td>
</tr>
<tr>
<td>Topic 5</td>
<td>.86</td>
<td>.53</td>
<td>.23</td>
<td>.31</td>
<td>1.09</td>
</tr>
<tr>
<td>Topic 6</td>
<td>1.14</td>
<td>.76</td>
<td>.26</td>
<td>.29</td>
<td>1.40</td>
</tr>
<tr>
<td>Topic 7</td>
<td>1.23</td>
<td>.75</td>
<td>.22</td>
<td>.24</td>
<td>1.44</td>
</tr>
<tr>
<td>WARRANT</td>
<td></td>
<td></td>
<td></td>
<td>29.95***</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td>.41</td>
<td>.30</td>
<td>.08</td>
<td>.14</td>
<td>.48</td>
</tr>
<tr>
<td>Topic 2</td>
<td>.68</td>
<td>1.04</td>
<td>.30</td>
<td>.36</td>
<td>.97</td>
</tr>
<tr>
<td>Topic 3</td>
<td>.07</td>
<td>.21</td>
<td>.03</td>
<td>.11</td>
<td>.10</td>
</tr>
<tr>
<td>Topic 4</td>
<td>.09</td>
<td>.27</td>
<td>.08</td>
<td>.22</td>
<td>.17</td>
</tr>
<tr>
<td>Topic 5</td>
<td>.25</td>
<td>.33</td>
<td>.25</td>
<td>.33</td>
<td>.50</td>
</tr>
<tr>
<td>Topic 6</td>
<td>.50</td>
<td>.40</td>
<td>.11</td>
<td>.18</td>
<td>.61</td>
</tr>
<tr>
<td>Topic 7</td>
<td>.47</td>
<td>.35</td>
<td>.15</td>
<td>.20</td>
<td>.62</td>
</tr>
<tr>
<td>BACKING</td>
<td></td>
<td></td>
<td></td>
<td>11.63***</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td>.23</td>
<td>.33</td>
<td>.06</td>
<td>.11</td>
<td>.29</td>
</tr>
<tr>
<td>Topic 2</td>
<td>.49</td>
<td>.70</td>
<td>.14</td>
<td>.29</td>
<td>.64</td>
</tr>
<tr>
<td>Topic 3</td>
<td>.11</td>
<td>.33</td>
<td>.16</td>
<td>.36</td>
<td>.27</td>
</tr>
<tr>
<td>Topic 4</td>
<td>.11</td>
<td>.24</td>
<td>.18</td>
<td>.36</td>
<td>.28</td>
</tr>
<tr>
<td>Topic 5</td>
<td>.02</td>
<td>.08</td>
<td>.10</td>
<td>.19</td>
<td>.12</td>
</tr>
<tr>
<td>Topic 6</td>
<td>.05</td>
<td>.11</td>
<td>.09</td>
<td>.16</td>
<td>.14</td>
</tr>
<tr>
<td>Topic 7</td>
<td>.38</td>
<td>.39</td>
<td>.32</td>
<td>.41</td>
<td>.69</td>
</tr>
<tr>
<td>REBUTTAL</td>
<td></td>
<td></td>
<td></td>
<td>6.06***</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td>.01</td>
<td>.06</td>
<td>.02</td>
<td>.06</td>
<td>.03</td>
</tr>
<tr>
<td>Topic 2</td>
<td>.03</td>
<td>.15</td>
<td>.01</td>
<td>.06</td>
<td>.04</td>
</tr>
<tr>
<td>Topic 3</td>
<td>.03</td>
<td>.13</td>
<td>.11</td>
<td>.44</td>
<td>.15</td>
</tr>
<tr>
<td>Topic 4</td>
<td>.03</td>
<td>.11</td>
<td>.16</td>
<td>.32</td>
<td>.19</td>
</tr>
<tr>
<td>Topic 5</td>
<td>.05</td>
<td>.11</td>
<td>.15</td>
<td>.26</td>
<td>.20</td>
</tr>
<tr>
<td>Topic 6</td>
<td>.10</td>
<td>.22</td>
<td>.13</td>
<td>.22</td>
<td>.23</td>
</tr>
<tr>
<td>Topic 7</td>
<td>.18</td>
<td>.29</td>
<td>.11</td>
<td>.19</td>
<td>.29</td>
</tr>
</tbody>
</table>

Note: *\( p<0.1 \), **\( p<0.01 \), ***\( p<0.001 \); N=74.

Each statement generated by students was further categorized into two levels of claim, warrant, backing and rebuttal arguments in order to reveal its quality. The mean frequency of two levels of arguments generated by each group in each question shows a growing pattern overall, regardless of levels of claim, warrant, backing and rebuttal argument (Table 3). Repeated measures of ANOVA showed that an increase of level 1 arguments was found to be statistically significant when comparing earlier topics with later topics in all aspects, regardless of claim, warrant, backing and rebuttal (\( F(\text{claim})=53.50, p <0.0001; F(\text{warrant})=23.88, p <0.0001; F(\text{backing})=14.59, p <0.0001; F(\text{rebuttal})=3.03, p <0.005 \)). The increase of level 2 arguments was also found to be statistically significant when comparing earlier topics with
later topics in all aspects, regardless of claim, warrant, backing and rebuttal ($F_{\text{claim}}=9.38$, $p<0.0001$; $F_{\text{warrant}}=11.26$, $p<0.0001$; $F_{\text{backing}}=5.62$, $p<0.0001$; $F_{\text{rebuttal}}=7.89$, $p<0.005$).

**Conceptual Change**

The nature of each argument was judged and classified into three categories as correct, partially correct, and incorrect. The results show that the mean score of correct conceptions for each argument generated by each student increased from pre- to post-argumentation questions across all 7 topics, and 6 topics reached a statistically significant difference level ($T_{\text{topic 1}}=2.25$, $p=0.027$; $T_{\text{topic 3}}=3.31$, $p=0.001$; $T_{\text{topic 4}}=4.16$, $p=0.000$; $T_{\text{topic 5}}=7.18$, $p=0.000$; $T_{\text{topic 6}}=5.05$, $p=0.000$; $T_{\text{topic 7}}=3.96$, $p=0.000$). The mean score of partially correct conceptions decreased from pre- to post-argumentation for three topics and only one of the topics reached a statistically significant difference level ($T_{\text{topic 7}}=2.82$, $p=0.006$), and slightly increased for three topics. The mean frequency of incorrect conceptions decreased from pre- to post-argumentation for about six topics, and only one of the topics reached a statistically significant difference level ($T_{\text{topic 5}}=5.15$, $p=0.000$) (Table 4).

**Table 4.** Analysis of the correctness of conceptions for each argument generated at pre- and post-argumentation question by each student across seven topics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Pre-argumentation</th>
<th>Post-argumentation</th>
<th>Mean difference</th>
<th>T</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.66</td>
<td>.66</td>
<td>.89</td>
<td>.72</td>
<td>.23</td>
</tr>
<tr>
<td>PC</td>
<td>.45</td>
<td>.50</td>
<td>.32</td>
<td>.40</td>
<td>-.13</td>
</tr>
<tr>
<td>IC</td>
<td>.07</td>
<td>.18</td>
<td>.05</td>
<td>.20</td>
<td>-.02</td>
</tr>
<tr>
<td>Topic 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.74</td>
<td>.74</td>
<td>.97</td>
<td>.99</td>
<td>.23</td>
</tr>
<tr>
<td>PC</td>
<td>.64</td>
<td>.75</td>
<td>.81</td>
<td>.82</td>
<td>.18</td>
</tr>
<tr>
<td>IC</td>
<td>.04</td>
<td>.20</td>
<td>.01</td>
<td>.12</td>
<td>-.03</td>
</tr>
<tr>
<td>Topic 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.50</td>
<td>1.49</td>
<td>2.72</td>
<td>3.07</td>
<td>1.22</td>
</tr>
<tr>
<td>PC</td>
<td>.36</td>
<td>.73</td>
<td>.45</td>
<td>1.17</td>
<td>.08</td>
</tr>
<tr>
<td>IC</td>
<td>.20</td>
<td>.55</td>
<td>.20</td>
<td>.60</td>
<td>.00</td>
</tr>
<tr>
<td>Topic 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.32</td>
<td>1.27</td>
<td>2.00</td>
<td>1.39</td>
<td>.68</td>
</tr>
<tr>
<td>PC</td>
<td>.58</td>
<td>.91</td>
<td>.57</td>
<td>.76</td>
<td>-.01</td>
</tr>
<tr>
<td>IC</td>
<td>.12</td>
<td>.44</td>
<td>.05</td>
<td>.23</td>
<td>-.07</td>
</tr>
<tr>
<td>Topic 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.79</td>
<td>.80</td>
<td>1.56</td>
<td>1.09</td>
<td>.77</td>
</tr>
<tr>
<td>PC</td>
<td>.47</td>
<td>.60</td>
<td>.33</td>
<td>.46</td>
<td>-.14</td>
</tr>
<tr>
<td>IC</td>
<td>.69</td>
<td>.81</td>
<td>.19</td>
<td>.40</td>
<td>-.50</td>
</tr>
<tr>
<td>Topic 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.49</td>
<td>.93</td>
<td>2.28</td>
<td>1.52</td>
<td>.79</td>
</tr>
<tr>
<td>PC</td>
<td>.36</td>
<td>.47</td>
<td>.37</td>
<td>.48</td>
<td>.01</td>
</tr>
<tr>
<td>IC</td>
<td>.12</td>
<td>.23</td>
<td>.08</td>
<td>.22</td>
<td>-.04</td>
</tr>
<tr>
<td>Topic 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.59</td>
<td>.96</td>
<td>2.17</td>
<td>1.23</td>
<td>.57</td>
</tr>
<tr>
<td>PC</td>
<td>.88</td>
<td>.60</td>
<td>1.13</td>
<td>.85</td>
<td>-.77</td>
</tr>
<tr>
<td>IC</td>
<td>.23</td>
<td>.29</td>
<td>.16</td>
<td>.26</td>
<td>-.07</td>
</tr>
</tbody>
</table>

Note: C: correct conceptions; PC: partial correct conceptions; IC: incorrect conceptions; N=74.

**Conclusions and Discussions**

This study reports a Recurrent On-Line Synchronous Scientific Argumentation learning program that was developed based on the conceptual change and scientific argumentation theories in order to promote 8th grade students’ conceptual change and scientific argumentation ability in a physical science context. This study is a major step from
previous Web-based instructional learning programs, as it brings well-developed conceptual change and scientific argumentation pedagogy theories and models into Recurrent On-Line Synchronous Scientific Argumentation learning. In addition, our learning environment contains two layers of template to provide students’ guidance and support in constructing a good argument. It also provides the advantage of real time argumentation, so students can receive prompt rebuttals to their arguments. This helps to make learning more effective.

The results of this study are quite positive as they demonstrate that On-Line Argumentation learning is far more effective than conventional instruction for promoting students’ conceptual change and scientific argumentation. Our results add positive documentation to the current research that students who receive a Web-based learning course can perform better than a conventional group’s students in their physical science concept construction. In addition, we argue that most computer-assisted learning studies cannot effectively change students’ alternative conceptions or science learning because their instructional materials are not developed based on solid theories or models of conceptual change or science learning (She & Lee, 2008; Liao & She, 2009; Yeh & She, 2010). This study supports the idea that including argumentation and conceptual change theories into the design of Recurrent On-Line Synchronous Scientific Argumentation learning are important for success.

In addition, the results of the on-line scientific argumentation process indicated that the amount of arguments generated by student is significantly greater when comparing later topics with earlier topics, regardless of level 1 or level 2. It clearly demonstrates that students’ ability to generate arguments indeed increased from topic 1 to topic 7 across the semester. Moreover, the mean frequency of arguments increased significantly from earlier topics to later topics, regardless of claim, warrant, backing, and rebuttal. These results demonstrated that our design indeed improves the quantity and quality of argumentation in the physical science context, regardless of the data collected from tests or on-line scientific argumentation process.

Our results demonstrate that students’ argumentation significantly improves across a semester-long intervention of argumentation in a physical science context. The breakthrough we made fully supports Osborne et al.’s suggestions that a recurrent opportunity for students to be involved in argumentation may make difference (Osborne et al., 2004). Our results also confirmed that constructing a good argument is not a simple task and that students need guidance and support to help them scaffold and build their sense of an effective argument. With the supports of writing frames, which indeed cultivate their ability to generate claim, warrant and backing arguments fairly quick except rebuttal. This clearly supports the claim that the on-line synchronous argumentation learning environment, with the support of writing frames, indeed speeds up students’ ability to generate better and higher level arguments within very short period and continuously grow till the end.

Moreover, our data found that students’ ability to generate argumentation is not quite stable across a semester. It is rather difficult to generate rebuttal arguments and takes a much longer period to cultivate compared to the claim and warrant arguments. Our data indicated that the quality and quantity of claim and warrant arguments become stable after topic 5, and rebuttal arguments increased gradually all the way from topic 1 through topic 7. It indicated the need to provide students with recurrent opportunities to use argumentation in science in order to stabilize their ability of using argumentation and increase their level of argumentation.

The mean frequency of correct conceptions for arguments generated by each student increased from pre- to post-argumentation questions across all 7 topics, and 6 topics reached a statistically significant difference level. The finding is quite positive and promising with regards to changing students’ alternative conceptions from both the assessment and on-line argumentation process. It is clear that the success in conceptual change is due to the design of Recurrent On-Line Synchronous Scientific Argumentation learning, which organizes conceptual change into a series of pre-, experiment-related, and post-argumentation activities. Our results demonstrate that our design of taking conceptual change ideas is critical for successfully changing students’ conceptions through a series of argumentation activities, specifically the ideas of creating dissonance, which provide an opportunity for them to experience the dissonance and become further dissatisfied with their prior conceptions (She, 2004; Hewson & Hewson, 1988); providing plausible mental structures for students to reconstruct more scientific conceptions (She, 2004); and actively engaging students in the group discussions to shape their knowledge construction process and changing conceptions (Driver, 1995; Venville & Treagust, 1998).

Finally, the regression results show that the best single predictor for post-PSDAT sores was the pre-PSCT, followed by pre-PSDAT scores. It demonstrated that constructing arguments of quality will be restricted and hampered
without the sources of specific knowledge or relevant scientific evidence. It implies that students who hold less alternative conceptions are more likely to perform better in the post-Physical Science Dependent Argumentation Test.

Acknowledgements

This research is based in work supported and funded by the National Science Council (NSC 95-2522-S-009-001-MY3), Taiwan, R.O.C.

References


Analyzing the Learning Process of an Online Role-Playing Discussion Activity

Huei-Tse Hou
Graduate Institute of Engineering, National Taiwan University of Science and Technology, Taiwan // hthou@mail.ntust.edu.tw

ABSTRACT
Instructional activities based on online discussion strategies have gained prevalence in recent years. Within this context, a crucial research topic is to design innovative and appropriate online discussion strategies that assist learners in attaining a deeper level of interaction and higher cognitive skills. By analyzing the process of online discussion in depth using a specific instructional strategy, we may discover the characteristics and limitations of this strategy. This case study utilizes an online discussion activity adopting a role-playing strategy in a college course and conducts an empirical analysis to explore and evaluate both the content structure and behavioral patterns in the discussion process. We propose and adopt a new method of multi-dimensional process analysis that integrates both content and sequential analysis, whereby the dimension of interaction and cognition are analyzed simultaneously. Furthermore, we discuss the patterns, characteristics, and limitations of the role-playing discussions and provide suggestions as references for teachers who utilize online role-playing discussion activities.

Keywords
Online discussion, Role-playing, Behavioral pattern, Instructional strategy

Introduction

Instructional Activity Utilizing Online Role-playing Discussion

Online discussion instructional activities have been widely applied in higher education courses. Furthermore, their effectiveness has been extensively discussed by several researchers (e.g., Gilbert & Dabbagh, 2005; Hou et al., 2007; Yeh, 2010). Instructional activities based on online discussions and appropriate teaching strategies may include the development of the learner’s argumentation skills (Driver et al., 2000; Oh & Jonassen, 2007). This approach may help learners gain a deeper understanding and develop higher cognitive skills. As a result, a crucial research topic relates to designing customized, innovative online discussion strategies that facilitate teaching and allow learners to reach a deeper level of knowledge construction and develop advanced cognitive skills. The quality of online discussions is often influenced by the design of the underlying interactive mechanisms (Gilbert & Dabbagh, 2005). There are several interactive learning strategies that are used in discussion-based online teaching activities, such as peer assessment (e.g., Hou et al., 2007) and problem solving (Hou et al., 2008; Oh & Jonassen, 2007). By analyzing the process of online discussion within these strategies, the characteristics and limitations of each strategy can be identified because teachers may not be able to predict the ideal timing for intervention during online discussions (Mazzolini & Maddison, 2007). The findings of the process analysis may help teachers choose better strategies, intervene in discussions at more opportune times, and develop improved facilitative mechanisms that address known limitations. Many behavioral analyses of online discussion activities (Hou, 2010; Hou et al., 2007, 2008; Jeong, 2003) provide insight into process limitations when students conduct discussions without teacher guidance. They also offer teacher guidance strategies based on behavioral patterns such as interactive mechanism design and proper timing of guidance. Furthermore, developers of educational software may also refer to these findings to develop appropriate tools, such as designs with automatic behavioral analysis technology, which can be used to automatically detect behavioral patterns (Hou et al., 2010).

In addition to problem solving and peer-assessment interactive strategies, another teaching strategy involves asking students to learn through role-playing. This approach has been increasingly researched in recent years (e.g., Bos & Shami, 2006; Hou, 2011; Wishart et al., 2007). This type of interactive learning is commonly used to help learners develop skills to handle group decision-making (Bos & Shami, 2006; Pata et al., 2005). Some studies suggest that role-playing also keeps learners motivated (Wishart et al., 2007) and improves communication skills among professionals (Chien et al., 2003). Furthermore, this strategy helps students develop abilities in problem solving by requiring them to assume different roles and confront unstructured problems in scenarios involving the professional domain of the given role.
Several studies explore the roles that students play during the process of collaborative learning (e.g., De Wever et al., 2008; Hara et al., 2000; Strijbos et al., 2004, etc.), such as source searcher, theoretician, summarizer and moderator (De Wever et al., 2008). However, these studies rarely focus on the cognitive process of online teaching activities in which students simulate real-life scenarios in assorted roles designated by teachers. The behavioral pattern of cognitive dimension and knowledge interaction in this role-playing learning process requires a more in-depth exploration. Role-playing should also be amenable to applications in discussion-based online teaching in which a teacher assigns roles and tasks to students (or students decide for themselves), and the students play the roles (such as requiring students to role-play different positions in a company) and discuss a given task such as solving the company’s management issues. The purpose of this approach is to increase student interactions during social knowledge construction and produce higher cognitive skills through realistic discussions. There are limited studies on the behavioral patterns of discussion-based online teaching involving role-play tasks. Hou (2011) examined the cognitive dimension of online role-play discussion activities, and concluded that students are better at the diversity of the cognitive aspect when role-playing in problem-solving tasks/scenarios. However, this study focused only on cognitive dimension analysis, and not the in-depth research on knowledge interactive behavior between the roles.

Therefore, an important and interesting topic for research is to explore the characteristics and limitations of both the interaction and cognitive patterns in online role-playing discussion activities. The research presented in this paper comprises an empirical case study exploring the use of role-play in discussion-based online learning in a higher education course. To analyze both the content of the discussion and the interactive behavior of its participants from multiple dimensions, we propose an approach that integrates the dimensions of interaction and cognition.

**A Multi-dimensional Process Analysis of the Interactions and Cognition of Online Discussions**

Many studies have explored the process of computer-assisted collaborative learning and developed various analytical methods and frameworks (e.g., Daradoumis et al., 2006; Hou et al., 2008). Furthermore, the analytical methods for online learning processes have gradually shifted towards the integration of multiple and diverse trends, such as the layered framework for evaluating online collaborative learning interactions proposed by Daradoumis et al. (2006). This framework includes a multidimensional analysis of social interaction and learning achievement. Hou et al. (2008) also explored the behavioral patterns in online teaching and learning activities that involve knowledge construction and the problem-solving process. Numerous studies have analyzed the process of online discussion as a teaching tool (e.g., Hou et al., 2008; Jeong, 2003). Furthermore, several studies have explored the frequency of interaction in online learning (Black et al., 2008), or conducted social network analysis to understand the state of the learners’ social interaction (e.g., Zhu, 2006). However, when focusing merely on the analysis of discussion frequency or social interaction status of online learning activities, the exploration on the “content of interaction” of the discussion is limited. To address this issue, an analysis of the content of discussions may help us gain a deeper understanding of a given discussion activity and the behavior of its participants. Quantitative content analysis has been used in several studies to explore the online discussion process (e.g., Gunawardena et al., 1997; Jeong, 2003). In those studies in which the messages of a learning community’s discussion were coded, the collected data allows for further analysis that can increase the understanding of the content structure of the entire discussion. However, simply understanding the content of discussions does not reveal much about behavioral patterns and does not allow us to understand the process of “content-related interactions”. To explore both content structure and behavioral patterns, we have therefore integrated content analysis and the lag sequential analysis (Bakeman & Gottman, 1997) as a way to infer sequential behavioral patterns based on discussion content. During the overall discussion process, sequential analysis allows us to determine whether students would further conduct a certain cognitive aspect of discussion after completing another cognitive aspect of discussion. Several studies apply sequential analysis to behavioral analysis regarding teaching and learning (e.g., Hou et al., 2007, 2009; Jeong, 2003; Sung et al., 2008). However, most of these studies only involve sequential analysis of behavioral patterns using a single coding scheme, and they rarely involve a sequential analysis in which additional dimensions are coded simultaneously. Using more dimensional schemes may better demonstrate behavior sequences when the discussion is influenced by multiple factors or may enable improved interpretation of the characteristics and limitations of the given discussion strategy.

Furthermore, sequential analyses of online discussions that cover both “interactive” and “cognitive” dimensions simultaneously are limited. Most of the analyses that have focused on knowledge construction interaction tend to address the learning community’s “interactive” behaviors such as sharing, coordination, and joint knowledge construction (e.g., Gunawardena et al., 1997; Hou et al., 2007). This paper provides an in-depth exploration of the “cognitive” processes during discussions including the phases of remembering, understanding, and analyzing, as
described in the Cognitive Process dimensions of the revised Bloom’s Taxonomy (Anderson & Krathwohl, 2001). Furthermore, we believe that the integration of the interactive and cognitive dimensions may help us realize the limitations regarding cognition and knowledge interaction among learners.

Based on the above discussion, we have developed a method that can simultaneously analyze the content and behavioral patterns of online discussions. Furthermore, this method can account for both dimensions of knowledge interaction and cognitive process in the discussions, and it helps us better understand the characteristics and limitations of various teaching activities. We propose an innovative analytical approach, which is a multi-dimensional process analysis of online discussion teaching activities. This approach integrates content and sequential analysis while also considering both the dimensions of interaction and cognition. With this proposed method, we devised appropriate coding schemes for the dimensions of interaction and cognition in online discussions. Each posting was coded and quantitatively analyzed according to the coding schemes of each dimension. This made it possible to measure the proportion of each code in order to better understand the content structure of cognition and interaction in learners’ discussions. Furthermore, to understand the behavioral patterns of the learning community’s interactions and cognition, a sequential analysis was conducted based on the derived codes. The above analyses reveal multi-dimensional data that simultaneously cover interactions, cognition, content, and behaviors. The results of these analyses may allow a more valid triangulation for facilitating an increase in the scope of the process analysis of online discussion instruction.

We explore the following research questions: 1. What are the content structure and behavioral patterns of an online discussion teaching activity using a role-playing strategy?; and 2. What are the characteristics and limitations of both the interactive and cognitive dimensions of an online discussion teaching activity using a role-playing strategy? To summarize, the specific purposes we wish to achieve in our study are as follows:

1. To conduct a typical role-playing-based online discussion learning activity using a general forum tool as well as an empirical observation and quantitative content analysis in which we explore the content structures of the learners’ overall knowledge construction interactions and the development of their cognitive skills.
2. To perform a sequential analysis of the coded data so that we may infer the visualized sequential behavioral patterns within the discussions in both the interactive and cognitive dimensions. These results will be cross-examined to further explore the characteristics and limitations of students’ social knowledge construction interactions and cognitive skills in role-playing-based discussions. Based on these findings, we will discuss and propose suggestions as references for teachers adopting role-playing strategies.

Method

Participants

There were 70 participants, 34 males and 36 females, in the study. All participants were college students majoring in information management. This study was implemented in a course which focused on common organizational behavioral issues and case analyses of business organizations in the context of online technologies. The participants all had basic Internet and information capacities and knew how to access and use a forum.

Design

We first coded each post and its responses using two coding schemes (i.e., a knowledge construction interaction dimension and a cognition dimension). The coded data were then used in quantitative content and sequential analysis to discover the content structure and behavioral patterns based on the content of discussion. This approach differs from analysis using the records or frequency of online operations (e.g., Black et al., 2008; Hou et al., 2010). Our analytical approach based on quantitative content data allows us to explore the dimension of knowledge content in learners’ interactions and to systematically decipher the behaviors hidden in the discussions. In addition, our approach is also different from that of another study in which only the proportion of codes in a quantitative content analysis were manifested. Our method allows a series of calculations based on the matrix of behavioral transition frequency (i.e., frequency transition matrix, conditional probability, conditional probability matrix, and adjusted residuals table) in the lag sequential analysis (Bakeman & Gottman, 1997; Hou, 2010) and enables us to conduct a more structured and visualized analysis of behavioral patterns. Additionally, in our study we utilized two coding schemes that enabled us to analyze both interactive and cognitive dimensions. This approach not only allows us to
understand the content structures of interaction and cognition but also permits us to generate two content structures and sequential behavioral patterns for cross examination and interpretation. Subsequently, a qualitative content analysis is conducted on certain key findings to provide better evidence and clarification for behavioral patterns.

**Tools**

*Online Discussion Forum*

To avoid pop-up advertising windows when using forums provided by commercial web sites, we have developed a forum with basic interactive functions as the tool for observation and recording in the study. Our forum contains only basic functions, such as posting a thread, replying to a thread, and listing all threads, and can only be accessed by our participants. After logging in, participants see a list of all thread entries and may choose to post new threads, click on a thread link to read it, check responses to each thread, or respond to a thread. Because a thread may have multiple replies, the system lists the names of those who posted threads and replies so respondents can be identified.

**Coding Schemes**

To uncover the characteristics and limitations of the interactions in students’ online discussions and to analyze interactive social knowledge construction, we adopted the coding scheme of the “Interaction Analysis Model (IAM)”, proposed by Gunawardena et al. (1997). This model, which is depicted in Table 1, has been widely used in numerous studies for analyzing learners’ interactive social knowledge construction in online discussions (e.g., Hou et al., 2007, 2009; Jeong, 2003). Each phase of the IAM model represents an interactive behavior in social knowledge construction (such as information sharing, co-construction, or negotiation). After students’ discussion was coded by the IAM scheme, it will help to understand students’ knowledge interactive process of discussion. To explain more clearly the meaning and scope of each code item, we also provided a typical brief discussion example in Table 1 as an illustration.

**Table 1. Interaction Analysis Model (IAM) (Gunawardena et al., 1997)**

<table>
<thead>
<tr>
<th>Code</th>
<th>Phase Description</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Sharing / comparing of information</td>
<td>Statement of observation or opinion; statement of agreement between participants</td>
<td>“I found a website that is also discussing our current topic. The website address is…”</td>
</tr>
<tr>
<td>C2</td>
<td>Discovery and exploration of dissonance or inconsistency among participants</td>
<td>Identifying areas of disagreement; asking and answering questions to clarify disagreement</td>
<td>“I have a different opinion on Student S’s argument. Why would student S think that the same concept existing in both argument A and argument B is varied? This is different from my initial concept”</td>
</tr>
<tr>
<td>C3</td>
<td>Negotiation of meaning/co-construction of knowledge</td>
<td>Negotiating meanings of terms and negotiation of the relative weight to be used for various agreement</td>
<td>“I think the various principles mentioned in theory C need further understanding and discussion, and I personally believe the second principle is more important”</td>
</tr>
<tr>
<td>C4</td>
<td>Testing and modification of proposed synthesis or co-construction</td>
<td>Testing the proposed new knowledge against existing cognitive schema, personal experience or other sources</td>
<td>“I think the current plan proposed by everyone is different from my personal experience. The data in database D can support my viewpoint and provide a reference for any plan revisions”</td>
</tr>
<tr>
<td>C5</td>
<td>Agreement statement(s)/application of newly constructed meaning</td>
<td>Summarizing agreement and meta-cognitive statements that show new knowledge construction</td>
<td>“After contemplating various options for a new plan, the new proposed plan is as follows…”</td>
</tr>
<tr>
<td>C6</td>
<td>Others</td>
<td>Discussions irrelevant to learning topics.</td>
<td>“The episode of the drama tonight was amazing; the female character is gorgeous!”</td>
</tr>
</tbody>
</table>
To uncover the characteristics and limitations of cognitive behavior in interactions during the discussions, we adopted the Cognitive Process dimensions of the revised Bloom’s Taxonomy (Anderson & Krathwohl, 2001), which has been commonly used to distinguish the dimensions of cognition in numerous studies (Rovai et al., 2009; Valcke et al., 2009). As illustrated in Table 2, each code represents a cognitive aspect. This code classification will help us understand students’ cognitive process in online discussion activities. Furthermore, we have provided discussion examples to explain each code item in Table 2. Regarding the review and research of the quantitative content analysis (QCA) method, Rourke & Anderson (2004) suggested that researchers apply the existing and widely used coding schemes to improve the validity of the study. Because the above-mentioned schemes of knowledge construction and cognition have been widely applied or discussed in previous studies, they may positively influence the validity of quantitative content analyses.

### Procedures

We conducted a 20-day role-playing discussion activity for the 70 participants to observe how they collaborated in the discussion of a simulation scenario. The scenario concerned a company facing poor internal management. In the activity, the teacher assigned 70 different roles of employees to each of the 70 participants, i.e., 70 students role-played 70 different positions in a corporate office. For example, student A played the financial manager; student B, the sales representative; and student C, the general manager. The 70 participants were asked to discuss and comment on the organizational management-related issues faced by the company. Since all students were involved in playing different roles and interacted simultaneously with each other in the discussion activity, simulated real-life interactions within an organization were achieved. The target objective of the task was to have 70 employees within a corporate company face poor sales performance and management bottlenecks, and to have them rectify and reform issues such as controversy in performance appraisal system and digital organization planning. Therefore, under the guidance of managerial supervisors and the coordination between departments and interdepartmental communication, the task required all role-playing students to act according their job positions, and to draft a proposal for the aforementioned issues on company’s management reform.

To make the simulation more realistic, during the whole discussion the participants were asked to show labels stating their job titles and names in their postings. Students making a response comment could do so by posting a thread that others could read and post a reply. To avoid interference from the teacher’s subjective guidance and intervention, the
teacher was not allowed to provide any form (physical or virtual) of guidance or intervention before the discussion was concluded.

Data Analysis

We manually coded each message in the discussion in the following manner: each thread was treated as a unit, and the coder coded a thread and all its replies as a message list based on chronological order (the thread content itself is the first message; the first reply is the second message, and so on). There were 219 messages in the discussion. The average number of the learners’ postings was not very high. However, in role-playing activities, the number of discussions was relevant to the roles students played. The students who played manager positions did not need to search and share much data or make massive discussions of work details. They only needed to integrate and make a decision based on the collected data. Therefore, the number of postings from these students might be lower than other roles. The number of posted messages for each student differed with the differences of his role category and level of participation in this case study. The sum of discussion numbers was limited, but the students had sufficient discussions for the assigned task in the instructional activity.

Based on the two above-mentioned coding schemes, messages that covered the closest to a particular item within the two schemes were coded chronologically. Upon coding, each thread generated a coded dataset of each scheme, yielding 219 C-codes and 219 B-codes after the coding was completed by researchers with professional backgrounds in psychology. To ensure inter-rater consistency between coders, we gave approximately 70% of the discussion (160 entries of discussion messages) to another coder with the same background to code. The inter-rater reliability Kappa of the dimensions of social knowledge construction (C-code) and cognition (B-code) was 0.67 (p< .001) and 0.71 (p< .001), respectively. We calculated the distribution of various codes (including code C and code B) to understand the content structure of the discussion. The coded data was chronologically arranged for sequential analysis. The sequential analysis was carried out by conducting the statistical analysis of a series of behavioral transfer sequences from a coding item to another coding item. The analysis process included frequency transition tables, condition probability tables, expected value tables, and adjusted residuals tables (Bakeman & Gottman, 1997; Hou, 2010). Through these processes, we were able analyze each sequence in the matrix (e.g., sequence A-> B), and test if the continuity of individual sequences achieved statistically significance. The significance shown in specific sequences illustrated a behavioral transfer pattern in cognitive and knowledge interaction of the entire community observed. Finally, we also conducted qualitative content analysis in some of the behavioral phenomena observed and carried out in-depth discussions in the overall research findings.

Results and Discussion

Quantitative Content Analysis

The distribution of the codes of social knowledge construction and cognition are shown in Figure 1 and Figure 2, respectively. Because the four codes of C4, C5, B3, and B5 were not found in our study, they are therefore excluded from the figures.

Figure 1 indicates that, in terms of knowledge-construction-related interactions, C1 (sharing and comparing) has the highest percentage (87.67%). This finding suggests that in this discussion-based online instruction, the college students in this study often focus on knowledge sharing and comparison, or they may develop other knowledge construction phases (i.e., C2, C3, C4, and C5) based on C1. The percentage of off-topic discussions (C6: 0.91%) is extremely low, indicating that the level of concentration in knowledge construction interactions may be better achieved through the strategy of role-playing.

Diversity is rather limited in the dimensions of knowledge interaction beyond C1 (i.e., C2, C3, C4, and C5); among these, the percentage of C2 (7.31%) is slightly higher than that of C3 (4.11%), whereas C4 and C5 do not appear. However, C2, C3, C4, and C5 are the key factors in the process of argumentation (e.g., Erduran et al., 2004). The finding that C6 (0.91%) is extremely low is consistent with the idea that learners adopting role-playing strategies are better motivated in learning (e.g., Wishart et al., 2007).
Figure 2 indicates that the dimension of cognition in the discussion content is mostly dominated by B2 (81.74%), a finding that suggests that roughly 80% of the cognition process in discussions consisted of understanding (such as giving examples or explaining). Notably, B1 (Remembering) (5.94%), B4 (Analyzing) (5.94%), and B6 (Creating) (5.48%) show similar proportions, whereas B3 (Applying) and B5 (Evaluating) were not found in the discussion content. These results indicate that the structure of the students’ cognitive processes in a role-playing-based discussion consists of remembering, understanding, analyzing, and creating. However, because role-playing focuses on the training of students’ decision-making capabilities (Bos & Shami, 2006; Pata et al., 2005), knowledge may be applied in the decision-making process (B3) to form different plans before they can be evaluated (B5). In our study, however, these two types of discussion are absent, indicating that the teacher should be aware of this limitation and work on this process when conducting a similar activity.

Sequential Analysis

The data coded above underwent sequential analysis to analyze further the visualized sequential behavioral patterns of the role-playing discussion content. After calculating the frequency transition tables, the condition probability tables, and the expected value tables (Bakeman & Gottman, 1997), we derived the adjusted residuals tables for the
two coding schemes, as illustrated in Tables 3 and 4. The z-score value of each sequence was calculated to determine whether the continuity of each reached the level of significance. Each row indicates a starting discussion behavior, whereas each column indicates which discussion behavior follows; a z-value greater than +1.96 indicates that a sequence reaches the level of significance (p<0.05). Based on these results, we were able to infer transition diagrams for behaviors that reached the level of significance, as depicted in Figs. 3 and 4, where arrows indicate the direction of the sequences.

Table 3. Adjusted residuals table of the knowledge construction interaction

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C6</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.79</td>
<td>-5.13</td>
<td>-1.62</td>
<td>1.53</td>
</tr>
<tr>
<td>C2</td>
<td>-0.92</td>
<td>2.36</td>
<td>1.18</td>
<td>-0.46</td>
</tr>
<tr>
<td>C3</td>
<td>-0.04</td>
<td>0.43</td>
<td>-0.43</td>
<td>-0.3</td>
</tr>
<tr>
<td>C6</td>
<td>0.2</td>
<td>-0.44</td>
<td>-0.22</td>
<td>-0.16</td>
</tr>
</tbody>
</table>

*p<.05

Table 4. Adjusted residuals table of the dimension of cognition

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>B4</th>
<th>B6</th>
<th>B7</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>-0.69</td>
<td>-0.38</td>
<td>1.05</td>
<td>2.38*</td>
<td>-0.32</td>
</tr>
<tr>
<td>B2</td>
<td>1.01</td>
<td>1.95</td>
<td>-6.52</td>
<td>-5.95</td>
<td>1.54</td>
</tr>
<tr>
<td>B4</td>
<td>-0.82</td>
<td>-0.89</td>
<td>3.66*</td>
<td>1.94</td>
<td>-0.39</td>
</tr>
<tr>
<td>B6</td>
<td>1.24</td>
<td>-0.1</td>
<td>-0.55</td>
<td>-0.34</td>
<td>-0.28</td>
</tr>
<tr>
<td>B7</td>
<td>-0.33</td>
<td>0.2</td>
<td>-0.31</td>
<td>-0.19</td>
<td>-0.16</td>
</tr>
</tbody>
</table>

*p<.05

Figures 3 and 4 indicate that the only significant sequence for interactive knowledge construction was C2->C2, whereas the significant sequences for the dimension of cognition are B1->B6 and B4->B4. C2->C2 indicates that in this role-playing-based discussion activity, the students showed continuity in how they defined or discussed the various different comments from others, whereas B4->B4 also indicates that students showed a certain degree of continuity in their analysis of a given topic of discussion.
The sequences of C2->C2 and B4->B4 indicate that when the strategy of role-playing is utilized, student discussion may show a greater tendency to focus on the discussion of different comments and opinions (C2) as well as a stronger focus and a greater degree of continuity in the process of analysis in the dimension of cognition.

In addition, although B1 and B6 average only 6% of the overall discussion, B1->B6 indicates that in the process of discussion, students occasionally moved directly from remembering to creation (e.g., formulating new decisions). We see that although the content structure includes four cognition-related codes (i.e., remembering, understanding, analysis, and creation) and a certain level of continuity of analysis (B4->B4 behavioral pattern), B3 (Applying) and B5 (Evaluating) are absent, which suggests that a gradual advancement of the cognitive discussion phase (e.g., B1->B2, B2->B3, or B3->B4, etc.) does not occur in the discussion sequences. We also noted that some participants recalled specific shared information or comments directly from memory and went straight to the planning and decision-making aspects of creation (B1->B6) without discussion (e.g., B1->B2, B1->B4).

In order to further explain these findings, we focused on the qualitative content analysis of students’ discussion to better understand its context. In qualitative analysis we found that students in the role-playing process have a certain degree of understanding (B2) and analysis process (B4), but the proportion of its analysis level (B4) is still limited. This result is similar to the finding of previous studies (Hou, 2011). However, we also found that new ideas would appear in some of the discussion context without full understanding, application, analysis and evaluation. This is similar to the above B1->B6 finding, as is in the following excerpt of a student discussion:

Sales rep A (# S0113): Enterprise Resource Planning; ERP, Supply Chain Management; SCM, Customer Relationship Management; CRM, Knowledge Management; KM and other systems ... can increase efficiency. (The student then describes the individual function of these systems...):

IT personnel (# S0011): As an IT personnel, my viewpoint is that our company does not merely become digitized, but it should also take action and become mobilized. First, we start with “digitizing all stores”, which allows customers to enter the store and search information with the digital service platform, as well as provides market information to attain information transparency ... (the student then explains his new proposal ...)

Taking the above discussion as an example, the student S0113 who played the sales representative mentioned several systems and their functions that he believed could be used for digital organization based on his own knowledge. However, the IT personnel (S0011) directly addressed his digital organization proposal without even analyzing or assessing sufficiently the information provided by the sales representative or other students (such as contemplating the evaluation of S0113’s aforementioned systems or assessing their feasibility). Also, his new proposal did not specify the necessary steps to implement, nor did it conduct a feasibility assessment. This example reveals that students may jump to conclusions or decisions without undergoing a sufficient and complete cognitive process, or they may directly treat or quote online information as answers (e.g., Chang & McDaniel, 1995; Wallace & Kupperman, 1997).

On the other hand, we also explored the behavioral differences in the different categories of roles. The roles assigned to students were divided into two main categories, roles that involved taking on the managerial position with more responsibility and authority (such as department managers), and roles that involved taking subordinate positions (such as rookie sales representatives). We discovered in the qualitative analysis that students who play the managerial roles tend to give brief instructions, compile others’ opinions, or devise thinking and planning approaches for subordinate employees. Such role-playing behavior can help students themselves in planning and integrating abilities, and at the same time, motivate other members in data analysis. On the other hand, students who played the subordinate roles focused more on practical experience in knowledge sharing and data collecting, discussing the topics in details. While previous studies have identified the behavioral categories of students’ on-line collaborative learning process in role-playing (e.g., De Wever et al., 2008; Strijbos et al., 2004), the results of this study further clarify the characteristics of role-playing behavioral pattern in the activity of simulating real-life scenarios. The results show that a discussion activity specifying different simulated roles can help students achieve a certain degree of communication and cooperation, and may develop their communication skills (e.g., Chien et al., 2003). However, the task in this case study evidently demands extra effort from students to appropriately apply online resources in order to solve problems and evaluate each proposed proposal. Furthermore, in this study, students illustrated in the overall discussion an inadequacy in two cognitive skills: application and evaluation, and their analytical skills were...
also very limited. Teachers and software developers may use above findings to determine the types of intervention needed to facilitate discussions that would promote completion of the cognitive process.

**Conclusion and Suggestions**

In this study, we attempted to use a method of analysis that integrates content and sequential analysis to explore the characteristics and limitations of a role-playing-based online discussion activity for the learning community.

As for the characteristics of a role-playing-based online discussion activity, our process analysis and discussion indicate that the students in our role-playing-based activity demonstrated a certain degree of cognitive content structure in their discussions, a certain degree of analysis of different opinions, as well as behavioral patterns of sustained concentration. These findings may suggest that the strategy of role-playing motivates learners (e.g., Wishart et al., 2007) and may develop and improve some argumentation skills, such as comparing and analysis of different opinions to propose their claims (Driver et al., 2000).

As for the limitations of a role-playing-based online discussion activity, we have discovered that the cognitive dimension of the discussion lacked the development of B3 (Applying) and B5 (Evaluating), both of which comprise the decision-making skills valued in role-playing activities (e.g., Bos & Shami, 2006; Pata et al., 2005). This indicates that the diversity of social knowledge construction is restricted. The gradual advancement of the cognitive process was also limited. Furthermore, while our analysis showed continuity (B4->B4), analysis (B4) was not sequentially correlated with other cognitive processes (e.g., B1->B4, B2->B4, etc.). Some students even jumped directly from memorized knowledge to creation (B1->B6), indicating that they may jump to conclusions without going through a sufficient and complete cognitive process.

Based on these limitations and the above discussion, we propose the following suggestions for teachers when guiding learners in a role-playing-based discussion activity:

1. To help learners develop better cognitive skills, teachers may review the limitations of the discussions we discovered in the dimension of cognition when teacher intervention does not introduce and focus on promoting the depth of learners’ cognitive processes. For example, teachers may post messages to guide students to think about relevant applications (B3) (e.g., reminding the students that certain information gathered may be applied to solve a certain component of the issue of corporate reform or asking them to think about possible applications) or prompt them to evaluate different pieces of information and comments (B5: Evaluate) (e.g., reminding the learners to take note of the feasibility of certain plans and evaluate them) as a way to reinforce cognitive aspects that may neglected in the discussions. Teachers may also trigger connections between analysis (B4) and other cognitive aspects (e.g., triggering B2->B4, B3->B4, B4->B5) as a way to ensure a complete and in-depth cognitive process in the discussion.

2. To improve social knowledge construction, teachers may introduce more structured strategies to promote diversity in knowledge construction. For instance, they may divide the discussion activity into data-collecting, stating opinions, coordinating and reviewing each plan, and formulating decisions. These efforts increase the scope of discussions in a more structured and organized manner and promote interactive social knowledge skills such as negotiation (C3), the ability to apply past knowledge to the present, including reflection on and review of this practice (C4), and the capacity to organize creative thoughts generated by the group (C5).

Lastly, we have discovered that an analytical approach integrating interaction and cognition may allow in-depth analysis of the content structure and behavioral patterns of students’ online discussion process under a certain instructional strategy. This finding could be adopted as an evaluation method in future studies of online discussion instructional strategies.

Moreover, one worthwhile study for developers of intelligent discussion-based teaching systems is the design of an automated mechanism that integrates sequential analysis into online discussion or general learning platforms and automatically detects the learning process (e.g., Hou et al., 2010). In contrast with post-event, batched behavioral analysis, this approach allows researchers and teachers to evaluate instantly the behavioral patterns in online learning and to guide the learning community in a timely fashion.
In designing the discussion activity, we recommend that teachers design a series of scenarios that offer students the opportunity to change roles in many different tasks. This design may help enhance knowledge construction and diversity of cognitive thinking. Such an approach awaits future empirical researches for further in-depth analysis. In addition, there is much to be examined in the domain of role-playing-based online learning, including how realistic the learning community’s role-playing is, and how factors are correlated to learning motivation and learning performance.

Acknowledgement

This research was supported by the projects from the National Science Council, Republic of China, under contract number NSC-99-2511-S-011-007-MY3, NSC-98-2511-S-011-006, NSC-97-2631-S-003-002, and NSC-97-2511-S-011-004-MY3.

References


Hou, H. T., Chang, K. E., & Sung, Y. T. (2010). Applying lag sequential analysis to detect visual behavioral patterns of online


A Context-Aware Mobile Learning System for Supporting Cognitive Apprenticeships in Nursing Skills Training

Po-Han Wu1, Gwo-Jen Hwang2*, Liang-Hao Su3 and Yueh-Min Huang1

1Department of Engineering Science, National Cheng Kung University No.1, University Rd., Tainan city, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology No. 43, Sec.4, Keelung Rd., Taipei, Taiwan // 3Department of Information and Learning Technology, National University of Tainan No. 33, Sec. 2, Shulin St., Tainan city, Taiwan // cincowu@gmail.com // gjhwang.academic@gmail.com // verylikeapig@gmail.com // huang@mail.ncku.edu.tw

*Corresponding author

ABSTRACT

The aim of nursing education is to foster in students the competence of applying integrated knowledge with clinical skills to the application domains. In the traditional approach, in-class knowledge learning and clinical skills training are usually conducted separately, such that the students might not be able to integrate the knowledge and the skills in performing standard nursing procedures. Therefore, it is important to develop an integrated curriculum for teaching standard operating procedures in physical assessment courses. In this study, a context-aware mobile learning system is developed for nursing training courses. During the learning activities, each student is equipped with a mobile device; moreover, sensing devices are used to detect whether the student has conducted the operations on the correct location of the dummy patient’s body for assessing the physical status of the specified disease. The learning system not only guides individual students to perform each operation of the physical assessment procedure on dummy patients, but also provides instant feedback and supplementary materials to them if the operations or the operating sequence is incorrect. The experimental results show that the students’ learning outcomes are notably improved by utilizing the mobile learning system for nursing training.

Keywords

Mobile and ubiquitous learning, Sensing technology, Nursing education, Cognitive apprenticeship, Mastery learning theory

Background and Motivation

In traditional nursing education, apprenticeship is usually adopted. Michele (2008) further proposed cognitive apprenticeship teaching to conduct clinical nursing. The experimental results show that cognitive apprenticeship can help promote nursing skills and exploration as well as reflection during learning processes. In such learning activities, demonstration-providing, exercise-leading and feedback-giving are done by experienced nursing staff or experts; meanwhile, the abilities of the nursing students to work independently are evaluated (Woolley & Jarvis, 2007).

Although a one-on-one teaching mode brings better learning achievements for students, the actual number of teachers is usually insufficient to support this teaching mode. Instead, a one-to-many teaching approach is commonly used in actual teaching activities. Such an approach to learning often affects the students’ learning efficiency and effectiveness (Stalmeijer, Dolmans, Wolfhagen, & Scherpber, 2009). Moreover, in traditional classes, although dummy patients are used to help the students to learn to identify and collect the life signs of each body part, the effect is not satisfying since the operations of the students cannot be recorded and no instant feedback or supplementary materials can be provided.

To cope with these problems, some scholars have attempted to implement information technology in nursing activities. For example, Chang, Sheen, Chang, and Lee (2008) used online media as supplementary material for nursing learning. With an increasing demand for nursing professionals, how to promote the quality and effectiveness of nursing education has become an important issue. Facing this issue, researchers must take into account the design of teaching materials, tools and modes, and how to efficiently and effectively share nursing knowledge and practical experience to equip students with better abilities to deal with unexpected clinical situations (Guo, Chong, & Chang, 2007).

In order to adopt diverse learning methods to increase students’ learning motivation, in addition to online teaching, mobile devices, such as cell phones or Personal Digital Assistants (PDA), are also used in nursing activities (Mansour, Poyser, McGregor, Franklin, 1990; Young, Moore, Griffiths, Raine, Stewart, Cownie, & Frutos-Perez, 1994).
Using mobile devices as a learning tool for nursing education can provide abundant clinical teaching materials; in addition, the learning process of using mobile devices can be used as a basis for evaluating students' learning achievement (Dearnley, Haigh, & Fairhall, 2008; Hung, Lin, & Hwang, 2010; McKinney, & Karen, 2009). From related research, it has been found that mobile devices are often used as a knowledge acquisition tool in clinical nursing. Such a learning mode provides feasibility, but lacks functions of interaction and exercise. Hence, in the nursing programs that emphasize actual practice, how to establish a learning environment that provides personalized guidance and feedback for students to practice skills and apply knowledge in clinical situations is worth exploring.

Recently, due to the rapid development of sensing technology, combining real-world contexts with digital systems has become an important learning mode. Many researchers have attempted to combine sensing technology with mobile technology to build up context-aware ubiquitous learning, and have applied this technology to teaching activities in different subjects, such as natural science (Chiu, Tseng, Hwang, & Heller, 2010; Chu, Hwang, Tsai, & Tseng, 2010; Hung, Tsai, & Yang, 2008; Peng, Chuang, Hwang, Chu, Wu, & Huang, 2009), math (Zurita & Nussbaum, 2004), language (Chen & Chung, 2008) and social science learning (Hwang & Chang, 2011; Shih, Chuang, & Hwang, 2010). Some researchers further established digital libraries for supporting context-aware ubiquitous learning activities (Hwuang, Tsai, & Tseng, 2010). In this learning environment, the system can detect real-world situations via sensing technology, and guide students to learn through mobile devices in actual contexts (Uden, 2007; Hwang, Tsai, & Yang, 2008); the sensing equipment includes Bluetooth Technology (González-Castaño, García-Reinoso, Gil-Castañeira, Costa-Montenegro, & Pousada-Carballo, 2005) Radio Frequency Identification (RFID) (Hwang, Kuo, Yin, & Chuang, 2010) and Global Positioning Systems (GPS) (Huang, Lin, & Cheng, 2010). The major benefit of context-aware ubiquitous learning is to provide personalized scaffolding and support for students to observe and experience real-world situations so as to construct personal knowledge (Hwang, Yang, Tsai & Yang, 2009). The students, as a result of interaction with real contexts and learning systems, can conduct independent thinking and enhance their learning motivation to further promote learning achievement (Chu, Hwang, & Tsai, 2010).

Present research concerning context-aware ubiquitous learning is mostly outdoor ecological learning (Hwang, Tsai, & Yang, 2008; Hwang, Yang, Tsai, & Yang, 2009; Ng & Nicholas, 2009; Hwang, Kuo, Yin, & Chuang, 2010; Chu, Hwang, & Tsai, 2010). Skills training has not attracted scholars' attention until recently. For example, Hwang, Yang, Tsai and Yang (2009) developed a context-aware ubiquitous learning environment for guiding inexperienced researchers to practice single-crystal X-ray diffraction operations with step-by-step guidance and feedback. The experimental results showed that the context-aware ubiquitous learning mechanism is beneficial for cultivating students’ problem-solving abilities and operational skills.

In this study, we attempt to develop a nursing skills training system based on mobile and sensing technology for guiding students to practice the standard operating processes of respiratory assessment. Through a combination of the cognitive apprenticeship strategy and a context-aware ubiquitous learning environment, the students are not only provided with personalized guidance to strengthen their nursing skills, but are also offered prompt feedback and review to enhance their nursing knowledge.

Mobile System with Cognitive Apprenticeship Strategy for Physical Assessment

This study attempts to establish a nursing skills training system via mobile devices for students to learn the standard operating processes of physical assessment in a context-aware ubiquitous learning environment. The standard operating processes include collecting life sign information, physical assessment of patients, identifying diseases, and giving immediate nursing treatment. The framework of the learning system is shown in Figure 1.

The learning environment is a simulated sickroom, in which the dummy patients exhibiting physical symptoms are located. When the students approach a dummy patient (i.e., the learning target), the RFID reader on the mobile device detects the tag on the patient and provides relevant information, including the patient's name, symptoms (e.g., having a fever and having much sputum in the past week), and case history (e.g., having had a stroke five years ago). Afterward, the learning system guides the students to observe the dummy patient and collect data following the standard process of physical assessment. When the students finish the physical assessment procedure, the learning system immediately calculates their degree of mastery (DM), which represents the time needed to correctly complete
the physical assessment procedure in comparison with the expected completion time of an expert level learner. The DM of Student Si is calculated using the following formula (Barsuk, Ahya, Cohen, McGaghie, & Wayne, 2009; Block, 1971; Carroll, 1963):

\[
DM(Si) = \frac{\text{expected completion time}}{\text{student completion time}} \times 100\%
\]

Figure 1. System framework

The student collects and assesses the life sign of this body part.

Information of life sign:
(1) Blood pressure: 176/98 mmHg
(2) Temperature: 39°C
(3) Pulse: 110 times/min

Figure 2. Interface for detecting pathological signs
In this study, the focus of the instruction is the “respiratory system.” The student first logs onto the learning system and chooses a case from the scenario case database for practicing physical assessment. The student is guided through the standard operating process of physical assessment. In the “demonstration and guidance” stage, the student has to diagnose a dummy patient according to the case history offered by the system and the hints given by the sensing technology to answer questions. During the examination, the student gathers physiological information about the patient via an RFID reader on the mobile device. The system will give life signs corresponding to different positions. For example, after the student detects the RFID tag on the chest via the mobile device, the system will give information of breathing frequency (times/second). Other physiological information includes temperature, pulse and blood pressure, as shown in Figure 2. Through the life signs detection exercise, the student is able to familiarize him/herself with visual examination, palpation, percussion and auscultation, and determine the treatment of the patient according to the given information.

When the student conducts the standard operating process for physical assessment, the learning system compares the information retrieved from the case database with that provided by the student to ensure the correctness of each step. If the operation is incorrect, missing or overlooked, the system will give feedback to the student, as shown in Figure 3.

During the learning process, the student can view personal mistakes or missing steps, and repeat the exercise to reach a degree of mastery through "observation and reflection”, as shown in Figure 4.

For example, during the process of practicing palpation, the student detects vibration information caused by the voice of the dummy patient, breath movement and the position of the diaphragm, among which the tags of the left and right paths represent the patient’s expansion situation. In addition to palpation, visual examination information, percussion and auscultation are also collected based on the tags in the standard operating process. The final step of the standard operating process is to examine the patient’s blood test report, as shown in Figure 5. Afterward, the system presents some similar diseases for the student to identify and fill in according to the gathered symptoms.

After the student completes a practice of the physical assessment procedure, the learning system presents the current degree of mastery, as shown in Figure 6. For the first case of this illustrative example, the patient is suffering from pneumonia. Assuming that the student has spent 25 minutes correctly completing the physical assessment procedure...
and the expected completion time is 20 minutes, we have DM = \((20 ÷ 25) \times 100\% = 80\%\). Usually the teachers would define a higher standard for degree of mastery, such as 90%; therefore, the learning system will guide the student to spend more time practicing the procedure of checking the pneumonia case.

Figure 4. Interface for providing the correct steps

Figure 5. Interface for showing the patient’s blood test information
The student’s degree of mastery is shown in the table. The student needs to practice the cases marked in red.

Case 1 (pneumonia) and Case 2 (left pneumothorax) need to be practiced to reach a higher degree of mastery.

Figure 6. Interface for showing the student’s degree of familiarity with the standard operating process for the disease

As a result of repeated exercise, the student’s standard operating process skills gradually become immediate responses, reaching a degree of mastery. When the frequency of making mistakes reduces, the system will gradually lessen the hint-giving to assist the student to independently complete the physical assessment standard operating process. The student can thus achieve a degree of mastery through repeated practice. When the student correctly answers the same question three times consecutively, the system will determine and present the student’s degree of mastery based on his/her operation time. The flow chart for the mobile nursing training system is shown as Figure 7.

Experiment Design

The study aims to adopt cognitive apprenticeship teaching as a framework to train the students in learning physical assessment standard operating processes via a mastery mechanism in a context-aware mobile learning environment for mastering the procedures as experts.

Subjects

The subjects included two classes of fourth graders of the Nursing Department at a Science and Technology University in Kaohsiung County in Taiwan. A total of forty-six students voluntarily participated in the study. One class was assigned to be the experimental group and the other was the control group. The experimental group, including twenty-two students, was guided by the mobile supported system with cognitive apprenticeship to conduct physical assessment courses, while the control group with twenty-four students was guided by the traditional approach with learning sheets. All of the students were taught by the same instructor who had taught that particular nursing course for more than ten years. The standard respiratory system physical assessment operating process was constructed by two experienced teachers in the Nursing Department.
The student is trained following SOP.

The degree of mastery is shown.

The group questions are examined to see if they reach the mastery standard.

Mastery practice

The program is finished.

yes

The student logs onto the system.

All cases are verified for three successive times of correctness.

Randomly selected case question groups are given.

The case is removed from the database.

yes

The case is verified for three successive times of correctness.

The correct number of times of assessing cases is recorded.

Hints or supplementary materials about correct SOP are given.

yes

SOP of the student is verified for correctness.

The correct number of times of assessing cases is recorded.

Research Tools

The research tools in this study included learning achievement test sheets, mid-term exams (written and skill tests), questionnaires of learning attitude, questionnaires of cognitive load, and questionnaires for the acceptance of the mobile learning system. The test sheets were developed by two experienced teachers. The pre-test sheets consisted of two groups of questions about physical assessment (each group has four to six short-answer questions) and two short-answer questions about blood tests; the post-test sheets included ten multiple-choice questions (40%), seven multi-select questions (42%), and eight matching questions (18%). For the mid-term exam, the questions about the physical assessment of the respiratory system were extracted (42.5%). These questions included multiple choice (32.5%) and situational questions (10%). The skill tests evaluated the degree of accuracy (100%) and degree of smoothness (100%) of the actual operation.

The questionnaires of learning perceptions, cognitive load and reception of the mobile learning system were compiled by the researchers and revised by four experienced experts. Those questionnaires were presented using a six-point Likert scale, where “6” represented “strongly agree” and “1” represented “strongly disagree”.

The questionnaire of learning perceptions consisted of twelve items. Its Cronbach's alpha value was 0.925. The questionnaire of cognitive load consisted of four questions. Its Cronbach's alpha value was 0.897. The students in both groups were asked to complete the questionnaires after the learning activity.
The questionnaire for the acceptance of using the mobile learning system included two scales; that is, four items about “the ease of use of the mobile learning system” and three items about “the usefulness of the mobile learning system”. The Cronbach's alpha values for these two scales were 0.906 and 0.923, respectively; and the Cronbach's alpha of the entire questionnaire was 0.964.

**Experiment Procedures**

The flow chart of the experiment is shown in Figure 8. Before the experiment, the two groups of students took a two-week course about the basic knowledge of the respiratory system, which is a part of the formal nursing curriculum. After the course, a pre-test was conducted to evaluate the background knowledge of the two groups of students before participating in the learning activity.

In the beginning of the learning activity, the students in the experimental group first received a 30-minute instruction concerning the operation of the mobile learning system and the learning mission. Afterward, they were guided by the learning system to find each dummy patient and collect physical data for each case from the patient following the standard operating procedure. At the start, the learning system shows plenty of hints and supplementary materials to the students. After several practices, the system gradually reduces the amount of support to the students if they have achieved a higher DM.

On the other hand, the students in the control group learned with the traditional approach; that is, they were guided by the teaching assistant and were provided with a learning sheet, on which the learning missions and the situational questions were described. During the learning activity, the control group students were also guided to collect physical data from the dummy patients, for which printed instructions about the patients’ information (e.g., the patient's name, symptoms and case history) were provided. The students were asked to follow the instructions on the learning sheet to practice nursing operations and answer the questions; moreover, they could repeatedly practice the
standard operating process skills by themselves on the dummy patients after watching the demonstration of the teaching assistant.

The activity lasted for one hundred and eighty minutes. After the learning activity, the students received a post-test and the post-questionnaires for measuring their learning attitudes, cognitive load and their acceptance of the mobile learning system. From the results of the pre- and post-tests, the effectiveness of the guiding system for the physical assessment course in assisting the learning of the students could be evaluated. Moreover, through the feedback collected via the questionnaires, their learning attitudes, cognitive loads and the perceptions of the use of the mobile learning system could be analyzed. Some of the students were further interviewed. In the week after the learning activity, the students took a mid-term exam which included a written test and an operation test.

**Results and Discussion**

The study proposes a guidance system for the standard operating process of a physical assessment course based on the cognitive apprenticeship approach, and examines the effect of such a model on the learning achievements of the students in the experiment. In this section, the experimental results are discussed in terms of the dimensions of learning achievement, learning attitude, cognitive load, and reception of the mobile learning system.

**Analysis of Learning Achievements**

The study aims to examine the effectiveness of the mobile system using standard operating processes based on the cognitive apprenticeship approach for improving the learning achievement of the students. The mean value and standard deviation of the pre-test scores are 53.50 and 10.43 for the control group, and 70.14 and 12.71 for the experimental group. According to the \( t \)-test result (\( t=4.87, p=0.00<.05 \)), a significant difference is found between the two groups. It is evident that the two groups of students did not have equivalent abilities prior to taking this unit, as shown in Table 1. In order to explore the effectiveness of the mobile system using standard operating processes based on the cognitive apprenticeship approach for improving the learning achievement of the students, an analysis of covariance (ANCOVA) is used to exclude this difference between the prior knowledge of the two groups.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>22</td>
<td>70.14</td>
<td>12.71</td>
<td>4.87*</td>
</tr>
<tr>
<td>Control group</td>
<td>24</td>
<td>53.50</td>
<td>10.43</td>
<td></td>
</tr>
</tbody>
</table>

\( *p<.05 \)

Table 2 summarizes the results of the post-test scores. The mean value and standard deviation of the post-test scores are 53.13 and 8.48 for the control group, and 78.14 and 9.4 for the experimental group. The analysis of covariance (ANCOVA) is used to test the difference between the two groups by using the pre-test scores as covariate and the post-test scores as dependent variables. The adjusted mean value and standard error of the post-test scores are 54.89 and 1.98 for the control group, and 76.2 and 2.09 for the experimental group. According to the results (\( F=45.26, p=0.00<.05 \)), a significant difference exists between the two groups. After excluding the influence of the pre-test results, the post-test results of the mobile system for the physical assessment course showed a significant difference. That is, the mobile learning approach had significant and positive effects on the learning achievements of the students for the physical assessment course.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Adjusted Mean</th>
<th>Std.Error.</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>22</td>
<td>78.14</td>
<td>9.40</td>
<td>76.2</td>
<td>2.09</td>
<td>45.26</td>
</tr>
<tr>
<td>Control group</td>
<td>24</td>
<td>53.13</td>
<td>8.48</td>
<td>54.89</td>
<td>1.98</td>
<td></td>
</tr>
</tbody>
</table>

\( *p<.05 \)

Table 3 presents the \( t \)-test results of the scores of the skills test. It is noticeable that the scores of accuracy and smoothness are 87.32 and 87.75 for the experimental group, and 77.83 and 75.21 for the control group. According to
the \( t \)-test result, it is found that the level of accuracy (\( t=2.20, p=0.03<0.05 \)) and the level of smoothness (\( t=2.41, p=0.02<0.05 \)) show a significant difference between the two groups. Hence, the mobile system with the standard operating process for the physical assessment course was effective in improving the learning achievements of the students.

<table>
<thead>
<tr>
<th>Skills test</th>
<th>Group</th>
<th>Mean</th>
<th>S.D.</th>
<th>N</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>experimental</td>
<td>87.32</td>
<td>10.26</td>
<td>22</td>
<td>2.20*</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>77.83</td>
<td>17.72</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>experimental</td>
<td>87.75</td>
<td>10.55</td>
<td>22</td>
<td>2.41*</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>75.21</td>
<td>22.38</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

\( *p<.05 \)

### Analysis of Learning Log

The learning log of this study is analyzed and presented as well. The analysis of the learning log can further explain that the mobile system with the cognitive apprenticeship strategy for physical assessment helps improve the students’ learning efficiency. The standard physical assessment procedure consists of 134 operations. In the traditional approach, the students are asked to complete the procedure (i.e., the 134 operations) within three hours. Through the mastery mechanism of the mobile learning design with instant feedback, the number of times a student can practice for each case is far more than that of a student who learns with the traditional approach, as shown in Figure 9. Owing to the personalized learning guidance and the instant feedback provided by the mobile learning system, it is found that the frequency of practicing the standard operating procedure for physical assessment was significantly increased for each case, implying that the students had more practice opportunities. For example, the students in the experimental group performed about 445 operations (i.e., 3.32 practices) on average for the first case within three hours, while those in the control group only performed 134 operations (i.e., 1 practice) within the same time.

![Figure 9. Mean of the students’ average number of practices](image)

### Analysis of Learning Attitude

The study analyzed the learning situations of the students guided by the mobile system with standard operating processes for the physical assessment course based on the cognitive apprenticeship approach and guided by the learning portfolios in order to understand the learning preferences and learning attitudes of the students in the experiment. As shown in Table 4, both of the scores that exceed four reveal that the two groups gave higher evaluation for self-learning. The mean scores for learning attitude are 5.23 for the experimental group and 4.58 for
the control group. It is noticeable that a significant difference exists between the two groups ($t=4.14, p=0.00<0.05$). Accordingly, the mobile system proposed by the study better assists the students in understanding, and in evaluating themselves.

<table>
<thead>
<tr>
<th>Table 4. $t$-test result of learning attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning attitude</td>
</tr>
<tr>
<td>experimental group</td>
</tr>
<tr>
<td>control group</td>
</tr>
</tbody>
</table>

$p<.05$

From the scores of the skills test, the levels of accuracy and smoothness of the experimental group are better than those of the control group. This suggests that the control group was overoptimistic concerning its self-evaluation, leading to a longer time for completion of the standard operating process. The overall procedure of the control group is correct, but lack of familiarity results in a longer thinking time. In contrast, the levels of accuracy and smoothness of the experimental group are higher because the mobile learning system provides practice of the actual operating process to enable the students to digest the standard operating process for the evaluation of the respiratory system and to concentrate on disease assessment.

### Analysis of Cognitive Load

The cognitive load questionnaire aimed at exploring the cognitive load situation faced by the students. Table 5 presents an analysis of the students’ cognitive loads. The means are 3.43 for the experimental group and 4.33 for the control group. The $t$-test result ($t=-2.74, p=0.01<0.05$) shows a significant difference between the two groups. It is evident that the mobile learning system of the physical assessment course created no additional cognitive load for the experimental group students. It should also be noticed that nursing courses include not only instruction of nursing knowledge, but also a large proportion of internship programs. From the results of the analysis, the cognitive load of the control group is comparatively high. Therefore, using the mobile learning system as a teaching strategy can lower the trainee nurses’ cognitive load while learning.

<table>
<thead>
<tr>
<th>Table 5. $t$-test result of cognitive load</th>
</tr>
</thead>
<tbody>
<tr>
<td>cognitive load</td>
</tr>
<tr>
<td>experimental group</td>
</tr>
<tr>
<td>control group</td>
</tr>
</tbody>
</table>

$p<.05$

### Analysis of the Acceptance of the Mobile Learning System

The analysis of the acceptance of the mobile learning system is shown in Table 6, from which it was found that most of the students gave positive evaluations. The four questions with a mean exceeding five are “It is easy to operate the PDA interfaces of the mobile learning system”, “It is easy to read the information on the PDA screens of the mobile learning system”, “The response speed of the mobile learning system is well-matched with the learning progress on the site” and “I think that PDA operation of the personal learning system is easy”. This represents that the students evaluated the ease of use of the system positively. It was also observed during the experiment that the students, after being taught once or twice, were familiar with the PDA operations and learning contexts. However, from the standard deviation of the first item (i.e., 1.02), it was found that not all of the students highly accepted the interface of the PDA system, although most of them gave positive feedback (the mean is 5.23). This finding implies that more careful design of the user interface is needed in order to facilitate more students, in particular, those who have little experience in using computers or mobile devices.

For “Usefulness”, the questions with a mean exceeding five include “The personal progress provided by the mobile learning system can benefit my learning achievement”, “The guidance of the mobile learning system is quite clear and effectively assists me to understand the learning content and steps”, and “Combining the mobile learning system and the real-world contexts is helpful to learning”. This represents the positive evaluation given by the students.

<table>
<thead>
<tr>
<th>Table 6. Degree of the acceptance to the mobile learning system</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>experimental group</td>
</tr>
<tr>
<td>control group</td>
</tr>
</tbody>
</table>

$p<.05$
<table>
<thead>
<tr>
<th>Scale</th>
<th>Questionnaire item</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easiness</td>
<td>1. It is easy to operate the PDA interfaces of the mobile learning system.</td>
<td>5.23</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>2. It is easy to read the information on the PDA screens of the mobile learning system.</td>
<td>5.45</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>3. The response speed of the mobile learning system is well-matched with the learning progress on the site.</td>
<td>5.41</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>4. I think that PDA operation of the personal learning system is easy.</td>
<td>5.27</td>
<td>0.70</td>
</tr>
<tr>
<td>Usefulness</td>
<td>5. The personal progress provided by the mobile learning system can benefit my learning achievement.</td>
<td>5.36</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>6. The guidance of the mobile learning system is quite clear and effectively assists me to understand the learning content and steps.</td>
<td>5.36</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>7. Combining the mobile learning system and the real-world contexts is helpful to learning.</td>
<td>5.23</td>
<td>0.61</td>
</tr>
</tbody>
</table>

**Discussion and Conclusions**

This study proposes a context-aware ubiquitous mastery learning strategy using the cognitive apprenticeship approach applied to the standard operating process of a physical assessment course that included collecting patients’ life signs and physical assessment information, identifying diseases, and immediate nursing treatment. The students interacted with the learning system via mobile devices. Through repeated practice and feedback, the students learned visual examination, palpation, percussion, assessment and immediate nursing care of respiratory diseases. The experimental results showed that the learning achievement and learning attitude of the experimental students were significantly better than those of the control group students, revealing the potential of this innovative approach in nursing education. Moreover, students who used the mobile learning system showed lower cognitive load than those who learned with the traditional approach, implying that the personal guidance and immediate feedback mechanisms did ease the learning pressure of the students via showing them an effective way of learning.

Such a learning process conformed to the mastery learning theory proposed by Carroll (1963) that teachers, based on the learning targets, teach and assist students to repeatedly practice for a desired achievement of the final goal. In the past decades, the effectiveness of the mastery learning approach has been shown by many researchers (Lalley & Gentile, 2009; Horiuchi, Yaju, Koyo, Sakyu, & Nakayama, 2009; Johnson, Perry, & Shamir, 2010). Narciss and Huth (2006) showed that immediate feedback can significantly benefit students’ learning achievement and learning motivation. Zimmerman and Dibenedetto (2008) further emphasized the effectiveness of using the mastery learning approach in the classroom.

Although mastery learning theory has been recognized as being effective, it is difficult to carry out since the cost of providing individual students with a personalized tutor in the real-world training environment is too high. The use of the context-aware u-learning approach has overcome this problem. It enables the students to understand and construct complete concepts of respiratory disease assessment and to further master the assessment skills. Experimental results show that not only the knowledge level but also the skill level of the students was effectively promoted as a result of the nursing training using the mastery strategy in context-aware ubiquitous learning.

On the other hand, from the interview results, it was found that the students in the experimental group showed different degrees of acceptance of the PDA system interface, which could be due to their different experiences of using mobile equipment. Consequently, arranging more operational practice using mobile devices at the beginning of the learning activity is necessary; moreover, it is worth investigating a better way of developing a user interface for inexperienced learners.
Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 98-2511-S-024-007-MY3 and NSC 98-2631-S-024-001.

References


Exploring Non-traditional Learning Methods in Virtual and Real-world Environments

Rebeka Lukman and Majda Krajnc
University of Maribor, Maribor, Slovenia // rebeka.lukman@uni-mb.si // majda.krajnc@uni-mb.si

ABSTRACT
This paper identifies the commonalities and differences within non-traditional learning methods regarding virtual and real-world environments. The non-traditional learning methods in real-world have been introduced within the following courses: Process Balances, Process Calculation, and Process Synthesis, and within the virtual environment through the European funded Lifelong Learning Programme project at the University of Maribor. The results, based on qualitative research in both environments show the appropriateness of non-traditional learning methods in comparison with traditional ones, although collaborative learning in both environments causes several frustration based on conflicts (personal or disagreements during the learning phase), influencing the efficiency of the learning process. This presents opportunities for improving and overcome emerging barriers by fostering motivation and interactivity.

Keywords
Non-traditional learning methods, collaborative learning, Virtual- and Real-world environments

Introduction
The traditional learning environment has established a prevailing type of the educational process within the higher education area, for many decades. Over the last 20 years, computer-based learning has gradually revolutionized and revitalized the university sector, becoming an icon of the 21st century higher education provision (Selwyn, 2007). The primary reason has been the possibilities for technologies to expand opportunities regarding communication, interaction, and collaboration (Harasim et al., 1995).

The information-age and faster life-styles have spurred changes in traditional educational approaches, which had become unsuitable due to extensive curriculums. Lecturers simply do not find enough time to explain all the materials, to solve case studies on the blackboard and to teach students skills, such as oral and written communication, critical and creative thinking, group work, and a usage of new learning technologies (Krajnc, 2009). The development of information-communication technologies (ICTs) has added a new dimension to the learning process and made virtual-learning a significant learning option. Virtual-learning presents an alternative for students and helps them find a balance between private life, career, and further education. It is one of the most dynamic and enriching forms of learning (Paik et al., 2004), reducing dependency on space and time. On the other hand, it offers both individual learning experiences, and opportunities to work together (Peat, 2000).

Knowledge and skills for innovative and complex problem solving within multicultural and multidisciplinary environments can be acquired through non-traditional learning methods which encompass a wide-concept of learning through problem solving, and collaborative and virtual learning. Collaborative learning has been found to significantly enhance the learning process and student participation (Benbunan-Fich and Hiltz, 2003). Furthermore, studies show that collaborative learning results in higher achievement and information retention than competitive or individual learning (Hooper & Hannafin, 1988; Johnson & Johnson, 1996).

Nowadays, virtual learning environments supported by ICT are widespread in higher education (McGill and Hobbs, 2008) and collaborative learning can be used, not only during classes or when doing homework, but also within a virtual environment. Collaborative learning becomes virtual collaborative learning when it takes place via computer communication systems (Bélanger, 2009), facilitating interaction among students, tutors, and experts in order to exchange information. Virtual collaborative learning is an emerging approach that embraces the characteristics of collaborative learning and computer-mediated communication in network environments (Choi and Kang, 2010).

The objective of this study was to explore non-traditional learning methods, such as collaborative learning and to identify their commonalities and differences in real and virtual environments. In both learning environments several sources of frustration or disappointments exist, affecting the motivation and engagement of students (Sáiz Borges, 2009). These possible frustrations can cause problems in several activities, such as critical thinking, knowledge
sharing, problem solving, and achieving the required learning objectives. Sometimes, all learning activities may be reduced and even halted due to various problems. The following questions were raised in order to investigate those problems arising within the collaborative learning process:

- What was the efficiency, level of the collaborative learning process within virtual and real-world teams?
- Have there been any obstacles/advantages in both the real-world and virtual environments, disturbing/spurring the achievements of the learning process?

Learning efficiency was evaluated from three perspectives: educational/cognitive (solving real-world problems creatively, acquainting new knowledge, skills, methods, etc.), time (required for the learning process), and sustainability and depth of knowledge acquired.

**Collaborative learning**

In contrast to traditional, lecture-based learning, collaborative learning is an interactive, group knowledge-building process (Harasim et al., 1995), presenting a form of learning and learner interaction (So and Brush, 2008). In particular, collaborative or group learning refers to instructional methods that encourage students to work together on academic tasks. Collaborative learning is fundamentally different from the traditional "direct-transfer" or "one-way knowledge transmission" model in which the instructor is the only source of knowledge or skills (Harasim et al., 1995). Knowledge is a consensus among members of a community of knowledgeable peers – something that people construct by discussing and reaching an agreement. Students re-acculturate – break from their previous communities of knowledge – by working collaboratively with other students to create new knowledge or transfer values, and skills (Bruffee, 1993), focusing on new solutions when confronted with real-world problems. Collaborative learning also represents a common phenomenon for achieving a required goal within an educational or working process, mobilizing all the participants, sharing their knowledge, experiences and skills. It acknowledges that each individual brings different information, ideas, values, capacities, perspectives, and historical experiences and approaches into learning (Felder et al., 2000).

**Collaborative learning within a real-world environment**

Collaborative learning within a real-world environment usually takes place in small groups of three or four students (Bullard and Felder, 2007a), composed of students with heterogeneous abilities and various degrees of knowledge. In such a team, students share their strengths and discover their weaknesses by learning from their knowledgeable and more experienced colleagues. After the team is formed, special rules for teamwork are introduced and accepted, including individual accountability, positive interdependence, face-to-face interaction, regular self-assessment of the team, and the development of teamwork skills (Felder et al., 2000; Bullard and Felder, 2007a). The lecturer also clearly defines the goals and outputs to be achieved, e.g. lab report or process design. Each team member is responsible for his/her part after having understood the tasks of other team members. Work is divided between individual and group activities, where emphasis is placed on interactive activities based on regular meetings. During these meetings, students develop interpersonal and teamwork skills such as leadership, communication, conflict-resolution, and time-management (Felder et al., 2000), organizing their time for solving problems and also doing homework outside the classroom.

**Collaborative learning within a virtual environment**

Online collaborative learning can be described within a context where the computer, information, and network technology facilitates interaction among learners for the acquisition or sharing of knowledge (Liaw et al., 2008). Some authors have discussed learning networks as groups of people who use CMC (computer – mediated communication) to learn together, at a time, place, and pace that best suits them and is appropriate to the task (Harasim et al., 1995). Collaborative learning within a virtual environment is composed of students from different countries (including different time-zones), nationalities and backgrounds, without any prior history of learning together. Virtual environment represents the only opportunity for collaboration between course participants, including the exchange of ideas, thoughts, knowledge and experiences, and the planning of activities, research results, and the final report. In such an environment, face-to-face interactions are not generally feasible (Franceschi
et al., 2000). Various course management tools (Blackboard, Moodle, Doodle) and e-mails supporting collaboration in the educational process are used (Wilczynski and Jennings, 2003) in order to ensure the successes of virtual teams. Collaborative learning environments (INVITE, C-VISions) and 3D environments (CLEV-R), which allow for asynchronous and group learning have been used over recent years (Monahan et al. 2008). According to Franceschi et al. (2000), synchronous tools supporting voice communication (Skype) can be considered a critical factor in enhancing group collaboration because voice adds a personal touch to the communication process. The distance between students is psychologically shortened and, therefore, they feel safer. However, collaboration within a virtual environment is mostly based on asynchronous tools, available to students at any time of their choice (Bates and Poole, 2003). Asynchronous communication may promote critical-thinking, but it may also cause irritation in relation to the gap in response times, the so-called login-lags (Benbunan-Fich and Hiltz, 1999). Login-lags are time periods in which text or voice messages from other individuals within a virtual learning process are not received. For example, the enhanced temporal flexibility of asynchronous communication leads to a significant increase in the time required to complete a discussion topic (Levin et al., 1990). Virtual discussion and information sharing might also be perceived as slow (Meyer, 2003) because non-native students might have some difficulties when communicating in a foreign language (usually English). Therefore, it is important that students express their ideas and thoughts very clearly using short sentences. Furthermore, technology can cause problems for students with poor IT equipment, faulty electronic supply or inadequate internet services (Morse, 2003) because they cannot follow virtual discussions or chats consistently. Individuals also need time to learn and build their confidence when working with online tools in order to become efficient (Scantlebury, 2009) and become familiar with the functions offered by online tools.

Education Methods at the University of Maribor

Traditional learning environment

At the University of Maribor (UM), Faculty of Chemistry and Chemical Engineering, real-world collaborative learning was incorporated into three courses: Process Synthesis, Process Balances, and Process Calculation. The lecturer of these three courses tried to include novel methods into the already existed educational process, promoting a shift from the lecturer-centred to student-centred approach. During the academic year 2004/5, 26 full-time students started to work in teams within the Process Synthesis Course in the third year of the vocational study programme. The average age of the students was 22 years. Until that time, they had not had any experiences working in teams. From the academic year 2005/6 until 2008/9, the number of students varied between 18 and 31. Each team was composed of 2 or 3 students and they selected the members of the team themselves. The lecturer agreed with their decision due the fact that it is easier to work when those in teams are friends who know and understand each other.

The content of the Process Synthesis Course was divided into 8 sections. After each section, students had to solve problems or complete tasks using a collaborative learning method, which had not been the prevailing type of learning at the UM, thus continuously broadening their knowledge. Each of them carefully read and analysed the assignment, including a real-world problem and were then obliged to solve it. Discussion took place after obtaining and comparing the results. The student with the more accurate results is invited to explain his/her method of problem solving to the others. Sometimes students divide a problem into sub-problems, especially when it is extensive, where they share and complete different tasks. E.g. one creates tables, another draws graphs, and the third carries out calculus problems, in order to obtain the final result. In addition to collaborative work in the classroom, students have to fulfil their homework assignments. Each of them solves his/her part of the homework and then explains it to the group members. The final report is usually written by one group member.

During the academic year 2006/7, collaborative learning was also integrated into Process Balances and Process Calculation courses in the second year. The average age of the students was 21 years and they also had not had any experience in collaborative learning. From the lecturer's perspective, the educational process in these two courses was less efficient in terms of balanced group work and individual assignments, as the one for Process Synthesis. The main reason was the higher number of students enrolled (between 30 and 45). It was difficult to manage the groups, since there was only one academic staff member (lecturer) and from this perspective the educational process was inhibited. The lecturer was sometimes incapable of efficiently helping all the teams since she had insufficient time to comment and answer all the questions. The results of Bullard and Felders’ (2007a; 2007b) study show that efficient collaborative learning requires a low student-academic staff ratio, in their case 14 students per academic staff, which is approximately a half of those attending the courses performed at the UM.
Virtual learning environment

In the academic year 2008/9, a virtual-learning process was introduced for the first time to students from the Faculty of Chemistry and Chemical Engineering at the UM. It was offered within the European Virtual Seminar (EVS) (EVS, 2009), which is part of the Virtual Campus for Sustainable Europe project, coordinated by the Open University in the Netherlands. During the academic years 2008/9 and 2009/10 students participated in the following courses, Decoupling of Environmental Pressures from Quality of Life (DEC), Communicating Strategies for Sustainable Development (COM), European Water Framework Directive, the Danube Basin (WATER), and in an Urban Solid Waste (WASTE) course.

The main goal of the EVS project is to foster an international, multidisciplinary dialogue on sustainable development among students from all over Europe, by using modern ICT. The main features of EVS are:

- A multicultural and multidisciplinary learning community of students and staff,
- A virtual-learning process supporting collaboration between geographically-dispersed students,
- A learning content consisting of current scientific, environmental, economic, and social problems,
- A virtual-learning technology based on modern ICT, facilitating collaboration, communication, and interaction between students and staff.

Within the virtual-learning process of EVS, there are no ‘normal’ lectures. The students have to work in international, multidisciplinary groups, composed of 5 to 7 individuals. Group members cannot organise face-to-face meetings, and the collaborative and social processes depend on communication using modern ICT. The learning process is spread over a relatively long period (i.e. 16 weeks), in order to create the best possible conditions for virtual collaborative learning and to allow the students to participate in EVS, alongside their regular study programmes. In order to support such an exceptional learning process, EVS courses are divided into five stages:

1. Orientation and go/no-go decisions to the next stage;
2. Group-forming and community building;
3. Group research proposal;
4. Case-study and group report
5. Rewriting the group reports and producing a policy summary (EVS, 2009).

Performance of virtual learning activities

During the academic year 2008/9, a focus has been given to the educational process within DEC virtual group. As mentioned in the previous section, all the courses consist of five stages. In the first stage (orientation) the main coordinator, together with tutors, form study groups, based on a diversity principle, where students from their own countries and with the same or similar backgrounds are arranged into different study groups. Then each student receives a student manual and a timetable. These activities present an introduction to the course. An individual student’s activity begins by filling-out the Pexpi – a Personal Expertise Page, including ‘personal data’ (name, gender, birthday), ‘about me’, ‘interests and hobbies’, ‘expectations of EVS’, ‘EVS availability’, ‘Expertise areas’, ‘Fields of interest’, ‘Learning and work experiences’, and ‘Suggestions’. As Rusman et al. (2009) claim, the main aim of the Pexpi is to build a relationship of trust and understanding between collaborating individuals. This Pexpi is essential before proceeding with further group and individual activities. Within the DEC groups, the second point of Pexpi entitled ‘about me’ has caused certain problems, because students were unsure as to, which information about themselves should be presented, that is the one regarding their study, private life or both. Most of the students wrote down information about their studies, character, and family members. During this phase, the preparedness for communication and collaboration by students emerged. Some of them wanted to share a lot of information, whilst others were more reserved. One of the students did not actually want to expose his personal information to the others and did not want to share this on Blackboard. Therefore, he sent the Pexpi to the main coordinator only.

In the second stage (group forming and community building), the actual collaboration begins. Each student group receives a case-study, with a short introduction to the topic. All the documents are accessible through Blackboard, which also presents the only opportunity for communication and collaboration. The learning process at this stage is divided into three activities of which one is individual:
First group activity: is to form a sustainable development definition, as related to the case study. The idea is that each student participates by giving his/her opinion about the definition in written form on Blackboard, followed by a virtual discussion trying to find a consensus.

Second group activity: relates to an understanding of the case-study, where students define the goals of the learning process and case-study, and activities required to realize them.

Third (individual) activity: relates to a follow-up of individual virtual-learning processes.

Although the deadline for the first stage was very short (about two weeks) students did not want to expose themselves and start with the discussion. When approaching the deadline in the DEC group, all the activities narrowed down to only two students, who accomplished the assignments alone. Other students ‘came back to the virtual class a few hours before the submission and just confirmed their written assignment. The most common sentence was ‘I agree’.

The third stage was a group research proposal. This stage includes only one group activity – a group research proposal, consisting of research goals, methods, expected results, closely following the project application forms. Within this stage is important that students gain experience about the project’s proposals. It is a 5 week learning process. After 2 weeks DEC group had not started with their activities. A lack of motivation occurred, based on the idea, that the group would be unable to prepare a research proposal. One student, who was very experienced in EU project proposals, decided to prepare the whole proposal. Once it was finished motivation at last occurred and other students submitted their ideas, improvements, etc.

The fourth stage (case-study) is composed of two activities, an individual and a group. The individual activity comprises individual research activities as related to their home country, whilst the group activity represents merging all the individual data gathered and preparing a common discussion and conclusion. Within the DEC group problems regarding individual activities emerged. Some students were unavailable during these individual activities and understood them as ‘virtual vacations’, resulting in an extension of the submission deadline. Due the lack of discussion, one student collected the wrong data for her country. As a consequence, her data were not in line with the data from others, thus leading to difficulties when preparing discussion and conclusions. When preparing the report, the majority of the group did not want to accept their workload and chose one student as their informal leader (without her acknowledgment or confirmation) and sent her all their data. The typical shift of work and responsibilities by the group work was noted.

The last stage is rewriting the group report and producing a policy summary. This stage is closely linked to the 4th stage. Activities relate to a rewriting of the group report, according to the experts and tutors recommendations. This stage comprises two groups and one individual activity. The group activities concern reporting, policy recommendations, and a reflexive group report. The individual activity represents an individual learning record. During this stage, the learning process in the DEC group went well, because all the students had gained experiences from the previous stages and were familiar with their assignments and collaboration processes.

Analysis of used non-traditional learning methods

Non-traditional learning methods have been observed within the traditional learning environment since 2004, and within the virtual environment since 2008, based on a qualitative study. A survey was carried out, where students from both groups, i.e. real-world classroom and virtual-classroom, were questioned about their opinions regarding the efficiency of collaborative learning and the obstacles/advantages that emerged during both learning environments. The aim was to evaluate the quality of knowledge acquired, lifelong learning possibilities, time needed for the study process in comparison with classical teaching and learning (classical lectures – a frontal method, where the lecturer uses several tools (e.g. ppt presentations, graphs, tables, blackboard) and students are passive participants of the learning process) and the quality of team collaboration.

The survey was composed of six questions. One was of the essay type, whilst the others were multiple-choice questions. Among these, one required an explanation of the chosen answer. Table 1 shows questions and results of the survey for both classes.
Table 1. The survey questions and results

<table>
<thead>
<tr>
<th>1st question: Courses and lectures based on non-traditional learning methods were:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Real</td>
</tr>
<tr>
<td>Virtual</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd question: The non-traditional learning methods gave me:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Real</td>
</tr>
<tr>
<td>Virtual</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3rd question: The knowledge obtained with non-traditional learning methods was:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Real</td>
</tr>
<tr>
<td>Virtual</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4th question: With non-traditional learning I saved:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Real</td>
</tr>
<tr>
<td>Virtual</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5th question: How would you evaluate the collaboration among team members?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Real</td>
</tr>
<tr>
<td>Virtual</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6th question: Disharmony in the team appeared at the:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Real</td>
</tr>
<tr>
<td>Virtual</td>
</tr>
</tbody>
</table>

The questionnaires were completed by 34 full-time students of the real-world class at the end of their academic years as well as by 23 students participating in the EVS project at the end of their academic years. The first question related to the appropriateness of lectures based on non-traditional learning methods from the student perspective. All students of the real world class agreed that courses and lectures based on problem solving, e-learning, and collaborative learning were appropriate because they worked continuously on the topic. Such work allowed them to help each other and check their knowledge. They claimed that non-traditional learning methods brought diversity and dynamics into their learning, making the educational process more relaxing, interesting, and where lectures were not tiresome. Furthermore they claimed that it was easier to learn the theory at home. One student argued: “I learned how to cooperate in a team. It was good experience for the future”. The other said: “I acquired the theory section after section and I did not need to learn much more at home.” Another replied: “I remembered a lot of theory in such a way. The learning was easier. It stayed longer in my memory.”

In the virtual-classes, 91% of students felt that non-traditional teaching methods were appropriate because they added a new dimension and enabled creativity, development of new ideas, virtual brainstorming and flexibility to the learning process. Some their arguments in favour of virtual-learning were:

“...it meets my way to learn. The learn effect of classical lectures is very low by me. I’m ‘approbated’ non-traditional learning student.” (#1 student)

The other students argued

“...it’s easier to organise the time. The work can be done better and faster, because working is done at the moment when some person is filling to do it...”, (#5 student)
or

“...it can give you more in depth knowledge about a specific subject through discussing it with people who have different perspectives. That can be a valuable experience in a future career.” (#9 student)

or

“...I am not able to participate in traditional learning methods, I work from 8-5 and learn in the evenings/weekend.” (#22 student)

The second question tried to ascertain the amount of knowledge acquired with non-traditional methods in comparison with the traditional. These results are based on students’ personal opinions and their perceptions of the acquired knowledge. The majority of students in the real-world classroom (88 %) felt that non-traditional teaching and learning methods gave them more knowledge than classical lectures. In the virtual classroom, only 39 % of students claimed that non-traditional teaching methods gave them more knowledge than classical ones, whilst 52 % indicated that non-traditional teaching methods gave them the same amount of knowledge as classical lectures, and 9 % thought that they acquired less knowledge by virtual-learning.

The third question tried to identify the quality of knowledge (integrity, depth, sustainability) acquired through non-traditional learning methods (virtual-learning and collaborative-learning) in comparison with classical lectures. It should be noted that these results are based on the personal opinions and perceptions of students. About two thirds of the real-class students claimed that knowledge acquired through non-traditional learning methods was holistic, deeper, and more sustainable and fostered real-world problem solving. One third of the real-class students said that the acquired knowledge was the same as that obtained through classical lectures. In the virtual-classroom, the majority (78 %) of students felt that knowledge acquired through non-traditional learning methods was holistic, deeper, and would last longer, whilst 17 % claimed that the quality of knowledge was the same as that obtained through classical lectures.

Students were also asked about the amount of time saved with non-traditional learning methods in comparison with classical learning. Two thirds (67 %) of the real class students said that they saved more time with non-traditional learning, 15 % claimed they spent less time, and the rest (18 %) said they spent the same amount of time. In the virtual-classroom, 22 % indicated that they had saved more time in comparison with classical lectures. Furthermore, 61 % of students felt that virtual-learning had been more time consuming, whilst 17 % said that both learning methods were equally time consuming.

The quality of the learning process in both the virtual and real-world environments depends on the collaboration among team members. Students were asked to estimate the amount of collaboration within their teams. More than half of the real-class students (71 %) felt that the collaboration was excellent, arguing that everyone could express their opinion, they understood each other and that, in general, the work was divided equally. Furthermore, they helped each other and were able to receive help from their colleagues if they did not understand a problem. On the other hand, 29 % of students merely said that collaboration was good. Their arguments for excellent collaboration were: “There were no misunderstandings in our team. We completed each other.” The other student claimed: “We learned how to be responsible and tolerant when working in team. Each told frankly his/her opinion”.

In the virtual class, the majority of students (91 %) describe virtual collaboration as good and 9 % as bad. No one affirmed that collaboration was excellent. Regarding this situation students’ arguments were:

“...Like in teams in classical lectures: Basically collaboration depends on the motivation and engagement of the whole group”, (#1 student)

“...The collaboration was not really bad I would say just normal. If you have no personal contact, then the commitment is lower, so other things that are more "real" have a higher priority...” (#9 student)

“...Sometimes we missed each other and it took a while to get an answer, but generally every team member got through an equal amount of work and we all collaborated very well.” (#17 student).
“...the distance and the lack of others' responsibility make the collaboration will not as ideal as it should be.” (119 student)

In the last question, disharmony within teams was exposed. Students had to focus on why and when the disharmony appeared. One half of the real-class students said that disharmony within the team appeared at the task or problem solving stage, 9% indicated that disharmony appeared at the personal stage when someone did not attend lectures or did not want to cooperate at problem solving. 41% of students did not answer the question. Probably within these teams misunderstandings or disharmony did not appear at all and for that reason they did not choose any of the stated answers.

In the virtual-class, 44% of students commented that disharmony in the team occurred at the task or problem definition stage. 28% believed that disharmony was a consequence of personal virtual relationships (individual conflicts within the group). The other 28% claimed that disharmony mostly occurred at the task or problem solving stage due to disagreements over research methods and workload or due to dissatisfaction with the results (Table 1).

Discussion

The outcome of this study about non-traditional learning methods at the UM suggests that these methods (collaborative work, virtual-learning) are appropriate for several reasons. One of them is that collaboration among students influences the obtaining of better learning achievements during the education process. These achievements are reflected through various learning activities, carried out within non-traditional learning methods. Thus, students have an opportunity to brainstorm, to clarify problems within the group, and to come up with new, innovative solutions. Another reason in favour of non-traditional learning methods is the saving of time dedicated to study. Most of the required workload is done within the group through collaboration, and there is no need for additional work at home. From the efficiency perspective, students also tend to learn quicker within a group, and gain several skills, such as how to learn, who to ask for help, from whom to learn or how to find useful information.

The results of the survey show that the real-world class students appreciated non-traditional learning methods. This conclusion is obvious from the results of the 1st, 2nd, 3rd and 4th questions. Many students (two thirds and above) selected the first choice of these questions, which meant positive opinions towards non-traditional methods. The majority said that collaboration among team members was excellent. This statement is understandable because they are young, desire changes in the study process, and to be successful. Such results could be attained only with understanding and harmony which was mostly present in the real-world teams. Students, who were incorporated into this new way of learning, fulfilled the whole course’s obligations (written exams, laboratory exercises) up to the summer holidays, which meant that they become acquainted with the theory through electronic assessment of knowledge and teamwork. No one expressed that collaboration was bad. This would happen when work was distributed unequally i.e. when most of the problems were solved by one or a few of them and the others were totally inactive, or when the other members did not attend lectures regularly.

In the virtual-teams, students claimed that knowledge acquired was more holistic, deeper, and would last longer in comparison to traditional learning methods. The main reasons for this are so-called “login-lags” and efficiently-used time at virtual-discussions. Students are forced to do the assignments alone and some of them also in advance, so they are prepared for virtual-discussions. From an individual point of view, many of the questions and doubts are solved before the answers from other students are received. Because of individual approaches to problem solving within a group, the knowledge acquired tends to be longer lasting and deeper. However, the real sustainability and integrity of knowledge acquired in both the virtual and the real-world environments will be reflected in the future – when these students occupy a certain position in society and accept holistic and sustainable decisions for real-world problems.

In the virtual-class, students favour collaborative learning, although they consider it more time-consuming. They have to become familiar with IT tools, which are sometimes very complex and require the basic knowledge of computer and media science not needed by the traditional learning methods, and thus consuming more time. Group collaboration was denoted as good (not excellent) due to conflicts emerging at the personal and problem-solving levels. The reason for “not excellent” collaboration within the virtual groups might be caused by the lack of
motivation, which occurred during the individual research activities, when there were fewer discussions, and the level of collaboration among students was lower. The consequence was that nobody knew what other participants were doing and how their research was progressing. Another problem of collaboration was that sometimes some participants disappeared (due to personal circumstances) for 3 weeks or more – it was impossible to reach them and this affected teamwork negatively. Therefore, collaboration in virtual-teams tends to be stressful because there is no face-to-face interaction between students, tutors and experts. Disharmony on individual and process levels reduces efficiency, consumes energy and provokes negative feelings, which influence the quality of future teamwork. Negative feelings are reported to the tutors in the form of emails. One of the students reported: “... I don’t know what the rest of the group thinks as I have not gotten any email reactions yet. Perhaps it is readable between the lines, but I’m getting increasingly frustrated by our progress as a group ... But in a group that is to my view extremely passive, comments are not reacted upon, no feedback is given on each other’s parts, even arranging a Skype meeting is hard and the forum is hardly used (even if specifically asked to).” At this stage a tutor being involvement was necessary.

The success of virtual-learning depends on mutual relationships, communication, and group motivation. All these facts are important because real-world contacts, emotions, body language, and voices are absent in the virtual-world. Within the virtual environment, students perceive others through their writing, but they do not know anything about their capabilities, personalities and other features. For many students it is difficult to function within the virtual-world. Negative emotions and the absence of will can slow down or even stop progress.

Conclusions and Outlook

This case study at the UM, Slovenia represents a first attempt at the university for evaluating non-traditional learning methods based on the reactions of students. This represents an important step for the university and its further performance using non-traditional learning methods, due to the fact that most of the lectures carried out at the university are still done in a traditional way (lecturer-centred). Therefore, this study could be useful for educators or system designers for gaining some insight into non-traditional learning methods, including the challenges and difficulties faced by both lecturers and students.

This study showed that there are certain commonalities and also some differences regarding the non-traditional learning methods, based on the students’ opinions. In both environments the students preferred non-traditional learning methods. The majority of the students thought that any knowledge obtained this way would be more holistic and sustainable. They claimed that any disagreements occurred at the task or problem solving stage, where students were dissatisfied with the research methods chosen or results obtained. It is important to point out that non-traditional learning methods require a full commitment by all group members. The success of a learning process is, besides group collaboration, based on personal involvement, motivation, and the passions of each participant regarding the learning process. The results showed differences from the amount of knowledge obtained perspective. Whilst the majority of students in the real-classes claimed that they gained more knowledge than with traditional learning methods, the virtual-class argued that the amount of knowledge acquired with non-traditional learning methods was the same as using traditional methods. Another difference reflects the time-dimension. Whilst the majority of the real-world classes claimed that with non-traditional learning methods they saved time for studying, the majority of the virtual-class claimed that non-traditional methods consumed more time, compared to the traditional. The reasons lie in the gaining of IT skills, as argued in the discussion section.

There is still room for improvements in non-traditional learning methods in order to overcome these obstacles. In real-world environment, more emphasis needs to be placed on bridging individual preferences and developing strong social networks within teams, which would positively influence teamwork. Therefore, collaborative learning should be organized in groups (3-5). Based on the author’s experiences, up to 20 students could participate in the collaborative learning real-world class in order to achieve an efficient learning process. It was observed that smaller groups created a more personal learning approach, which affected the educational process in a positive sense.

In the virtual-class, more efforts should be directed towards decreasing frustrations by improving motivation and interactivity. Motivation could be strengthening by creating a sense of community and by building trust between students. In order to achieve the latter, a written personal commitment is to be signed at the beginning of virtual
collaboration within EVS, thus preventing people from changing their attitudes later. Such a commitment would also include an agreement about the frequency of virtual-meetings, preventing “login-lags”. Furthermore, there is a need for a clear, direct, quick, and continuous (if possible within a 24-hour time-frame) communication within the virtual-environment, otherwise the efficiency of the learning process is reduced, as observed during academic years 2008/9 and 2009/10.

In our study, some students claimed that virtual-learning helped them to improve their English. However, this interpretation requires in-depth research of the students' English competences, including the presentation of statistical data concerning language competences before and after the virtual-learning activities. Thus, our future work will focus on multiple-point data, specifically regarding the efficiency of the virtual-learning process from the conflict-factor perspective, including the role of a tutor as mediator. However, collaborative learning, especially virtual, represents a future challenge and more research needs to be done in order to design a better virtual learning environment.

References


European Virtual Seminar on Sustainable Development, Open University of the Netherlands, School of Science (OUNL). Available at: http://www.ou.nl/evs (last accessed 16 August 2009).


Learning Achievement in Solving Word-Based Mathematical Questions through a Computer-Assisted Learning System

Tzu-Hua Huang, Yuan-Chen Liu and Hsiu-Chen Chang

Graduate School of Curriculum and Instruction, National Taipei University of Education, No.134, Sec. 2, Heping E. Rd., Da-an District, Taipei City 106, Taiwan (R.O.C.) // Graduate School of Communication and Technology, National Taipei University of Education, No.134, Sec. 2, Heping E. Rd., Da-an District, Taipei City 106, Taiwan (R.O.C.) // Longtan Elementary School, No.71, Yulong Rd., Jiaoxi Township, Yilan County 262, Taiwan (R.O.C.) // anteater1029@mail2000.com.tw // liu@tea.ntue.edu.tw // jane@ilc.edu.tw

ABSTRACT

This study developed a computer-assisted mathematical problem-solving system in the form of a network instruction website to help low-achieving second- and third-graders in mathematics with word-based addition and subtraction questions in Taiwan. According to Polyà’s problem-solving model, the system is designed to guide these low-achievers through the parts of the problem-solving process that they often ignore. The situations of verbal questions are visualised to walk the students through the course of thinking so they can solve the question with proper understanding of its meaning. We found that the mathematical problem-solving ability of experiment group students was significantly superior to that of control group students. Most of the participants were able to continue the practice of solving word-based mathematical questions, and their willingness to use the system was high. This indicates that the computer-assisted mathematical problem solving system can serve effectively as a tool for teachers engaged in remedial education.

Keywords

Low-achiever, Problem-Based Learning (PBL), Polyà’s Problem-Solving Process, Word-Based Mathematical Questions

Introduction

In traditional teaching, assessment of whether students had understood a mathematical problem was based on whether they could describe the correct arithmetic procedure. However, it was not enough to evaluate students’ mathematics concepts and abilities of solving math problems merely depending on their writing. Some oral interpretation and explanation should be considered from multiple assessment points of view (Hwang, Chen & Hsu. 2006). Several researches also pointed out that good problem solving skills are the key to acquiring a successful solution in learning mathematics (Gagne, 1985; Mayer, 1992). The cognitive thinking process, as well as abstract mathematical concepts and representations, is often extremely difficult for middle-school and elementary students. Many efforts have been made to explore alternative ways of teaching mathematics by creating curricula and didactic material incorporating new tools, pedagogical approaches, and models or methods, to engage learners in a more pleasant mathematical learning process (Szendrei, 1996).

Fleischner and Marzola (1988) indicated that the mathematics learning achievement of about 6% of all middle-school and elementary students were severely flawed. In one of Polyà’s speeches (1962), he expounded on learning, teaching and learning to teach. He stated the fact that only 1% of the students would need to study mathematics, 29% would need to use mathematics in the future, and 70% would never need knowledge of mathematics beyond the elementary level in their daily life (Mayer, 1992). Obviously, if a student could successfully solve a math problem by arithmetic calculation, that did not mean the student really understood it.

Problem-Based Learning (PBL)

Problem-based learning (PBL) is considered to be one of the most appropriate solutions to increase students’ learning motivation and to develop practical skills. It is a student-centered instructional strategy, and students can solve problems collaboratively and reflect on their experiences. It also can help students develop skills and acquire knowledge (Bruer, 1993; Williams, 1993). Polyà’s Mathematics Problem-Solving
Process is one of the PBL instructional strategies. Lee, Shen and Tsai (2008) used PBL instructional design to enhance students’ learning and they also specified that teachers can help students to regulate their learning.

Polya’s Mathematic Problem-Solving Process

The mathematician Polya was the first to introduce the concept of problem-solving model. Believing that mathematics is not all about the result, he argued that the essence of mathematics education lies in the thinking and creativity employed in the problem-solving process. In his book “How to Solve It” (Polya, 1945) he proposed the following four steps of problem solving:

Understanding the problem

Problem solvers must understand the meaning of a sentence; identify the known, the unknown and the relationship between them; and know what previously learned concepts are available for solving the problem.

Devising a plan

Problem solvers must clarify the relations between conditions in a question, utilise personal knowledge to develop ideas for solving a problem, and devise a plan.

Carrying out the plan

Following the planned path, problem solvers carry out various calculations and other required operations.

Looking back

Problem solvers examine the answer and carefully review the course that they went through in an attempt to see if this experience helps solve other problems or if other problem-solving paths exist.

There is no clear distinction between these four steps. If it is determined that the plan cannot be carried out, or the examination reveals errors, review must be conducted to see if every hint has been correctly understood. A new plan must then be devised for implementation. This is necessary to emphasise the fact that in problem solving the process is more important than the outcome.

Examples of Computer-Assisted Mathematics Problem Solving Tools

Most students view computer-assisted learning environments as aids in learning mathematics and as motivational tools (Lopez-Morteo & Lopez, 2007). However, previous computer-assisted problem-solving systems have incorporated all the problem-solving steps within a single stage, making it difficult to diagnose stages at which errors occurred when a student encounters difficulties and imposing a too-high cognitive load on students in their problem solving (Chang, K. E., Sung, Y. T., & Lin, S. F., 2006). Because the systems have not been investigated empirically, the effectiveness of applying the systems to actual teaching is unknown. With technological advancement and the arrival of the multimedia computer instruction era, the attention of more and more studies has been fastened on real-time, interactive teaching methods and mediums through multimedia computers. The use of computers to implement findings from qualitative research related to problem-solving teaching strategies can furnish more effective learning opportunities for learners.

Hour and Chen (1999) investigated effects of using personalised context examples in CAI (Computer-Assisted Instruction) for fourth graders when solving mathematical word problems. Students' attitude toward the mathematical CAI was also examined in this study. The main finding was that students had favourable attitudes towards mathematical CAI. Data from personal interviews indicated that students were motivated
by personalised context examples in the CAI sessions. Generally speaking, students of middle and low problem-solving ability had positive mathematical belief after their exposure to personalised context examples CAI.

Frith, Jaftha and Prince (2004) found that when students used interactive spreadsheet-based computer tutorials in a mathematical literacy course, it brought to the foreground theories relating to the role of computer technology as a mediator for learning of mathematics. Outcomes showed that simple definitions of disadvantage were inadequate to account for the poor performance of students in the lower quartile. Olkun, Altun and Smith (2005) investigated the possible impacts of computers on Turkish fourth-grade students’ geometry scores and further geometric learning. The study used a pretest–intervention–posttest experimental design. Results showed that students who did not have computers at home initially had lower geometry scores. The result suggested that at schools, it seemed more effective to integrate mathematical content and technology in a manner that enabled students to make playful mathematical discoveries.

Ashton, Beevers, Korabinski and Youngson (2006) indicated that in a mathematical examination on paper, partial credit was normally awarded for an answer that was not correct but nevertheless contained some of the correct working. Assessments on computers normally marked an incorrect answer wrong and awarded no marks. This can lead to discrepancies between marks awarded for the same examination given in the two different media. In light of the findings, developments to the assessment engine have been made and some questions redesigned for use in real automated examinations. The results were obtained as part of the Project for Assessments in Scotland using Information Technology (PASS-IT), a major collaborative programme involving the leading educational agencies in Scotland.

Based on Polya’s four problem-solving steps (understanding the problem, devising a plan, carrying out the plan and looking back), Ma and Wu (2000) designed a set of interesting, active learning materials for teaching. Research outcomes indicated both students’ learning interest and achievement had improved. Chang (2004) incorporated strategies such as key-point marking, diagram illustration and answer review in the problem-solving process and developed a process-oriented, computer-aided mathematics problem solving system. The system was applied mathematical questions (mainly elementary-level arithmetic computation) with fifth graders as the subjects of the empirical study. Results showed that the system was effective in enhancing low-achievers’ problem-solving ability.

Chang et al. (2006) proposed a computer-assisted system named MathCAL, the design of which was based on the following four problem-solving stages: (1) understanding the problem, (2) making a plan, (3) executing the plan and (4) reviewing the solution. A sample of 130 fifth-grade students (average 11 years old) completed a range of elementary school mathematical problems. The result showed that MathCAL was effective in improving the performance of students with lower problem solving ability. These assistances improved students’ problem-solving skills in each stage.

Summary of the discourse above reveals that computer-assisted mathematics-problem-solving systems have a positive impact on children’s problem-solving ability.

System design and Framework

The system design of this study is based on the mathematical problem-solving process proposed by Polya (1945). It guides the children to think and solve the mathematical questions by using tag questions. Graphical representation strategy used in addition and subtraction for lower achievement children proposed by Fuson & Willis (1989) was taken as a reference for main design of strategy guide in solving the mathematics problem. In the study of Fuson & Willis, they found that in the situation of addition and subtraction, when the semantic structure conflicts with the solving strategy to get the unknown value, students might feel the answer is hardly to get. Fuson & Willis developed a schematic drawing to assist the lower achievement students to solve the problem in addition and subtraction with single step. The questions would be divided into four types based on the schematic drawings: (1) PUT-TOGETHER, (2) CHANGE-GET-MORE, (3) CHANGE-GET-LESS, and (4) COMPARE. Refer to Figure 1.
Figure 1. Schematic drawings filled in for exemplars of each word problem type

Figure 2. Students’ Problem-Solving Guidance Flowchart
**Students’ Problem-Solving Guidance Framework and Design**

Students’ problem-solving guidance process is shown in Figure 2. A fixed-format interface was employed to guide students in the problem-solving process to reduce the system operation learning load of users. Students can follow the problem-solving guidance process provided by the system and solve word-based mathematical questions.

**Teacher and Manager’s Online Real-Time Query and Management**

The system allowed the teacher’s online real-time inquiry about students’ account number and password, as well as students’ learning progress as shown in Figure 3.

![Figure 3. Teacher and Manager’s Online Real-Time Query and Management Framework](image)

**The screenshots of problem-solving process**

Screenshots of students’ mathematics-problem-solving process are shown in Figures 4-8.

![Figure 4. The screenshot of understanding the question](image)
Figure 5. The screenshot of understanding the question by animation guidance

Figure 6. The screenshot of listing the equation

Figure 7. The screenshot of listing the equation by animation guidance
Research Method

The participants in this study were selected from six classes of second and third grades in two elementary schools in Yilan County, Taiwan. The selection criteria were listed as follows: (1) Children with low mathematics achievement were identified by the class advisor that they had basic learning ability, need remedial instruction and had the average scores in mathematics posterior 25% of their class in the first monthly exam. (2) The class advisor consulted to the parents to find out the children who were willing to take the project as the experimental group.

For the students in experimental group trained by the computer-assisted mathematic-problem-solving system after school, there were seventeen second graders and third graders with low-achieving selected by above two criteria. Eleven students with low-achieving in mathematics that parents couldn’t send and pick up were the control group.

To test the discrimination and difficulty of the test questions, a pilot study was held between first monthly exam and second monthly exam of the first semester in academic year for the students from nine classes of second and third grades in two elementary schools in Yilan County, Taiwan. The test problems were divided into two examination papers, A and B. Both papers were designed for the students of second and third grades respectively. The difference between these two papers was the number magnitude. Most of the number used in the paper for the third grade was three figures; however, the number used in the paper for the second grade was double figures. After finishing the test, we combined the papers of the second and third grades to analyze of split-half reliability, discrimination and difficulty. Since the aim of this study was the remedial instruction of children with low mathematics achievement, the test questions were made based on the examination paper for the second grade (i.e., the number was double figures).

This study employed independent-sample dual-factor covariance analysis. The dependent variable was the posttest score, the independent variables included the group and grade, and the control variable (covariance) was the pretest score. Exploration was first conducted to see if interaction was present between the group and the grade, and then to see if significant discrepancy was present between different factors in terms of mathematical-problem-solving ability enhancement after conclusion of the experiment. Through independent-sample dual-factor covariance analysis, this study then looked into the presence of interaction between the group and the grade. It explored the impact of different factors on learning achievement of the computer-assisted mathematic-problem-solving system and determines whether significant discrepancy in learning achievement was present before and after the experiment in order to clarify whether the computer-assisted mathematic-problem-solving system could serve as an effective tool for remedial education. Answers given by students to the “Computer-Assisted Mathematic-Problem-Solving System Questionnaire” were analysed in terms of percentage for understanding of students’ opinion of the computer-assisted mathematics-problem-solving system.
Results and Discussion

System Achievement Analysis

Through the approach of descriptive statistics, this study first explained the average pretest and posttest scores of participants in each group and grade as shown in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>Grade</th>
<th>Total</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>2</td>
<td>17</td>
<td>11.27</td>
<td>15.00</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>11.33</td>
<td>13.83</td>
</tr>
<tr>
<td>Control</td>
<td>2</td>
<td>11</td>
<td>12.00</td>
<td>11.67</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>15.20</td>
<td>14.80</td>
</tr>
</tbody>
</table>

An intra-group regression coefficient homogeneity examination confirmed the assumed homogeneity of the covariance regression coefficients, so the dual-factor covariance analysis was conducted. The abstract is shown in Table 2. The result was not significant (F value=3.664; p=.068). Data by grade shows: F value = 0.162; p=.691; this interaction is not significant either. Significant interaction was only found between the groups. Therefore, the experiment did not result in a significant discrepancy between the low-achieving second graders and third graders.

Table 2. Abstract of Dual-Factor Covariance Analysis

<table>
<thead>
<tr>
<th>Variance Source</th>
<th>Type III SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explainable</td>
<td>84.384</td>
<td>4</td>
<td>21.096</td>
<td>5.667</td>
<td>.003</td>
</tr>
<tr>
<td>Interaction</td>
<td>104.711</td>
<td>1</td>
<td>104.711</td>
<td>28.130</td>
<td>.000</td>
</tr>
<tr>
<td>Independent Variable (grade)</td>
<td>0.603</td>
<td>1</td>
<td>0.603</td>
<td>0.162</td>
<td>.691</td>
</tr>
<tr>
<td>Dependent Variable (group)</td>
<td>25.042</td>
<td>1</td>
<td>25.042</td>
<td>6.727</td>
<td>.016</td>
</tr>
<tr>
<td>Covariance</td>
<td>37.351</td>
<td>1</td>
<td>37.351</td>
<td>10.034</td>
<td>.004</td>
</tr>
<tr>
<td>Main Effect</td>
<td>13.640</td>
<td>1</td>
<td>13.640</td>
<td>3.664</td>
<td>.068</td>
</tr>
<tr>
<td>Error</td>
<td>85.616</td>
<td>23</td>
<td>3.722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted Total</td>
<td>5658.000</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R Square = .496 (adjusted R square = .409)

Table 3 shows post-experiment learning achievements of experimental-group and control-group students through independent-sample single-factor covariance analysis. The outcomes indicated: F value=9.455**; p=.005. The adjusted average of the experimental group was 14.987, and the standard deviation was .495; the adjusted average of the experimental group was 12.474, and the standard deviation was .623.

The significantly higher scores of experimental-group students compared to control-group students (Pairwise Comparisons, Table 4) indicate that the computer-assisted mathematics problem solving system developed for this study can help teachers with remedial education and help students with basic word-based arithmetic questions.

Table 3. Abstract of Single-Factor Covariance Analysis

<table>
<thead>
<tr>
<th>Variance Source</th>
<th>Type III SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explainable</td>
<td>70.738</td>
<td>2</td>
<td>35.369</td>
<td>8.908</td>
</tr>
<tr>
<td>Interaction</td>
<td>99.844</td>
<td>1</td>
<td>99.844</td>
<td>25.146</td>
</tr>
<tr>
<td>Covariance</td>
<td>55.765</td>
<td>1</td>
<td>55.765</td>
<td>14.045</td>
</tr>
<tr>
<td>Main Effect</td>
<td>37.541</td>
<td>1</td>
<td>37.541</td>
<td>9.455</td>
</tr>
<tr>
<td>Error</td>
<td>99.262</td>
<td>25</td>
<td>3.970</td>
<td></td>
</tr>
<tr>
<td>Adjusted Total</td>
<td>170.000</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R Square = .416 (adjusted R square = .396)
Table 4. Abstract of Pairwise Comparisons

<table>
<thead>
<tr>
<th>Group</th>
<th>Group</th>
<th>Mean Difference</th>
<th>Standard Deviation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>Control</td>
<td>2.513</td>
<td>.817</td>
<td>.005**</td>
</tr>
</tbody>
</table>

*Indicates p<0.01

**Analysis of Students’ Problem-Solving Process**

Problem-solving changes at each stage of the problem-solving process were analysed and are presented in Table 5.

Table 5. Analysis of Students’ Problem-Solving Process

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
<th>Total</th>
<th>Mean or Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No. of students completing the practice</td>
<td>14</td>
<td>82.4</td>
</tr>
<tr>
<td>2</td>
<td>Average number of questions practiced per person</td>
<td>1104</td>
<td>64.9</td>
</tr>
<tr>
<td>3</td>
<td>Average problem-solving time per question</td>
<td>65.1</td>
<td>3.83 (min.)</td>
</tr>
<tr>
<td>4</td>
<td>No. of calculation errors</td>
<td>256</td>
<td>15.06</td>
</tr>
<tr>
<td>5</td>
<td>No. of problem-solving errors</td>
<td>168</td>
<td>9.88</td>
</tr>
<tr>
<td>6</td>
<td>No. of requests for question understanding assistance</td>
<td>129</td>
<td>7.6</td>
</tr>
<tr>
<td>7</td>
<td>No. of requests for expression or calculation assistance</td>
<td>938</td>
<td>55.2</td>
</tr>
</tbody>
</table>

Fourteen experimental-group students completed the practice, and the other three finished more than 40 questions. The average number of questions completed reaches 64.9, the average problem-solving time per question is 3.83 minutes, and the average number of practices is 7. Thereby we can see students' willingness to use the tool for learning problem solving is very high.

The average number of calculation errors per person detected by the system is 15.06, and the average number of completed questions is 64.9, indicating the ratio of calculation errors is 23.2%. Data indicates the ratio by which the system detects student’s calculation errors is not high, yet the researcher’s observation reveals that the proportion of the students who will examine or verify the calculation on their own initiative is not very high.

The average number of problem-solving errors per person detected by the system is 9.88, the average number of completed questions is 64.9, and the number of requests for question understanding assistance is 7.6. Yet the number of requests for expression or calculation assistance is as high as 55.2, indicating most of the students is still more inclined to look for the answer directly.

**System Questionnaire Analysis**

To collect information concerning user’s attitude toward and opinion of the computer-assisted mathematic problem solving system, this study designed the “Computer-Assisted Mathematic-Problem-Solving System Questionnaire.” The outcomes are presented in Table 6 and Table 7.

Table 6. Computer-Assisted Mathematic-Problem-Solving System Questionnaire Analysis A

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
<th>Percentage of students who agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>You feel that the operational instruction given by the system you are using is very clear.</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>After the explanation, I know how to work on each stage of the practice.</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>When there is an error, I understand the error hint given by the computer and I know where the problem is.</td>
<td>82.4</td>
</tr>
<tr>
<td>4</td>
<td>I feel it is easier for me to reach the answer through the computer than through the test sheet.</td>
<td>100</td>
</tr>
</tbody>
</table>
I feel practicing applied math questions through the computer is quite interesting. 100

I enjoy solving applied math questions through the computer. 100

I feel computer-assisted practice helps me develop a better ability to solve applied math questions. 100

In the stage of “I know the meaning of the question”, I feel reading the instruction given by the computer helps me understand the question better. 100

In the stage of “I know the meaning of the question”, I feel the questions asked by the computer help me understand the key points of the question better. 70.6

In the stage of “I know how to list the expression and calculate”, I feel the illustrated explanations given by the computer help me figure out the way to solve the question. 100

In the stage of “I know how to list the expression and calculate”, I feel the calculation explanations given by the computer help me solve calculation problems. 94.1

In the stage of “I will check again”, I feel this step encourages me to go back to examine my calculation again. 82.4

In the stage of “I will check again”, I feel this step encourages me to go back to examine my expressions again. 100

The method provided by the computer inspires me to think if there are other approaches. 94.1

I know completion of an applied math question must go through the following 3 steps: (please list them in actual problem-solving order) (1) I know how to list the expression and calculate (2) I will check again (3) I know the meaning of the question 82.4

Table 7. Computer-Assisted Mathematic-Problem-Solving System Questionnaire Analysis

<table>
<thead>
<tr>
<th>No.</th>
<th>Content</th>
<th>I know the meaning of the question</th>
<th>I know how to list the expression and calculate</th>
<th>I will check again</th>
<th>All very simple</th>
<th>All very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>The simplest step is</td>
<td>5.9</td>
<td>23.5</td>
<td>5.9</td>
<td>47.1</td>
<td>17.6</td>
</tr>
<tr>
<td>17</td>
<td>The hardest step is</td>
<td>11.8</td>
<td>17.6</td>
<td>17.6</td>
<td>47.1</td>
<td>5.9</td>
</tr>
<tr>
<td>18</td>
<td>Favorite step</td>
<td>23.5</td>
<td>29.4</td>
<td>41.2</td>
<td>5.9</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>Least desired step</td>
<td>0</td>
<td>35.3</td>
<td>11.8</td>
<td>52.9</td>
<td>0</td>
</tr>
</tbody>
</table>

System use attitude (questions 1~3)

Table 6 shows the majority of users (82.4-100%) thought that system operation and instructions could be easily understood, and they had no problem grasping the error feedback.

System assistance cognition (questions 4~7)

Table 6 shows all users agreed that solving math questions through the computer was easier than through the test sheet. It’s unanimous that solving word-based math questions through the computer is both interesting and enjoyable.

Mathematic problem-solving process cognition

Table 6 shows most of the users (82.4%) became familiar with the required stages for solving math questions, because they were able to list the problem-solving steps in correct order.

Cognition of each stage of the problem-solving process (questions 8~14)

Table 6 shows all participants agreed that motion pictures provided by the system to help understand questions helped them grasp the meaning of the questions. Most of the participants (70.6%) agreed that the guiding
questions designed for the system helped them identify the key points of the math questions. Over 94.1% of the participants indicated that the expression- and calculation-guiding motion pictures furnished by the system helped them make out the right problem-solving direction and complete the calculation. Over 82.4% of the participants said the double-check stage of the system helped them verify the expressions and calculations. A great number of the participants (94.1%) agreed that the method provided by the computer inspired them to think that there may have been better problem-solving approaches.

Preference of each stage of the problem-solving process (questions 16–19)

Table 7 shows about half of the users (47.1%) felt all the stages were very simple; 17.6% of the users considered all word-based questions to have been very difficult. About a quarter of the users (23.5%) regard expression listing and calculation as the easiest, whereas 5.9% each consider the question understanding step and the double check step the easiest, respectively. When asked again which step was most difficult, 5.9% of the users still said all the steps were very difficult, whereas 17.6% each consider the expression listing and calculation step and the double check step the hardest, respectively. Meanwhile, 11.8% see the question understanding step most challenging. Comparison of the favorite steps reveals that 41.2% of the users prefer the double check step, and 29.4% enjoy the expression listing and calculation step the most. Therefore, we can see users acquire the greatest sense of achievement from successful completion of a question. The result revealed that 5.9% of the users enjoy all 3 steps very much and 23.5% indicated their favorite part is the question understanding stage.

Conclusion

The ability to solve basic word-based addition and subtraction questions is the foundation for developing the ability to tackle complicated questions. Taiwanese students who experienced mathematical learning difficulties became more and more frustrated when they reached the middle grades because the difficulty and range of mathematical courses intensified and expanded. Hwang, Chen, Dung and Yang (2007) also recommended that teachers design problem solving activities to improve students’ mathematical representation skills. The main purpose of this study, therefore, was to develop an Internet-based computer-assisted mathematics learning system to help low-achieving elementary students improve their ability to solve basic word-based addition and subtraction questions and enhances their willingness to continue the learning. Outcomes indicated that the computer-assisted mathematical learning system developed for this study can serve as a supplementary tool that helps teachers with remedial instruction and enhances the problem-solving ability of low achievers.

References


Patterns of Interaction and Participation in a Large Online Course: Strategies for Fostering Sustainable Discussion

Jiyeon Lee
Myongji University, Bangmok College of Basic Studies, Republic of Korea//jiyeonlee@mju.ac.kr

ABSTRACT
This case study investigated patterns of interaction and participation in a large online course. 88 Korean undergraduates participated in online fora during 2 weeks. It was found that there was a comparatively high portion of metacognitive interaction and higher phase of knowledge construction. However, it was found that students posted required number of messages at once around due date and selectively responded to peers’ initiations without turn-taking. There were few responses despite many readings of peers’ postings to meet the required number and type of posting. The results showed that the discussion question and evaluation criteria influenced patterns of interaction and participation, and the phase of knowledge construction. Online instructional strategies were suggested for successful online discussion, especially to achieve sustainable discussion.

Keywords
Online interaction, Asynchronous fora, Knowledge construction

Introduction
The use of online courses and web-based communication has been growing exponentially with the expansion of online technology in colleges and universities. According to the results of a survey administered by the National Center for Educational Statistics (NCES), 56 percent of all degree-granting higher education institutions offered distance education courses in 2000-2001. 12 percent of all institutions showed that they planned to offer a distance education course in next 3 years (Waits & Lewis, 2003). The results from the most recent survey supported by Alfred P. Sloan Foundation confirmed fast growth of online learning in higher education. Over 4.6 million students were taking at least one online course during the fall 2008 term and more than one in four higher education students now take at least one online course based on the results (Allen & Seaman, 2010). Accordingly, fora have been considered one of the important components of online course (Garrison, Anderson, & Archer, 2001; Marra, Moore, & Klimeczak, 2004) with its widespread use.

Online fora characterized by asynchronous, text-based, and many-to-many communication have provided learners with flexible time and place independent learning, and facilitate a learning community with higher level of thinking with much learner-learner interaction (Bodzin & Park, 2000; Moore & Mara, 2005; Tutty & Klein, 2008). Learners have opportunities to share their own experience with peers in a shared electronic place which contributes to forming a good relationship (Chai & Tan, 2009). Text-based asynchronous interaction facilitates critical reflection by threaded discussion and enough time to reflect, think, and search for information (De Wever, Schellens, Valcke, & Van Keer, 2006; Pena-Shaff & Nicholls, 2004). All contents of online interaction are permanently stored and thus it helps students reflect on the learning process (Li, 2004). Suitable scaffolding and easily following the flow of discussion in online forums result in collaborative learning (Kanuka & Anderson, 1998; Veerman & Veldhuis-Diermanse, 2001). Sharing experiences, negotiating the meaning, and exchanging of resources and perspectives contribute to facilitating collective knowledge construction (Gunawardena, Lowe, & Anderson, 1997; Kanuka & Anderson, 1998; Moore & Marra, 2005). Students practice debate skills and gain deeper understanding by interaction with peers as a result (Reiser & Dempsey, 2002).

Despite the educational potential of using asynchronous threaded discussion, studies still found that students spent too much time presenting their ideas to develop discussion (Hathorn & Ingram, 2002). They often produced discussions that lacked relevance, and participated less in a discussion (Islas, 2004). Students sometimes did not read peers’ postings and posted only to meet the required numbers (Bodzin & Park, 2000). Students often simply posted answers to the questions and relied heavily on the instructor’s feedback (Chai & Khine, 2006).

As a result, online fora could not unconditionally guarantee educational potentials for the instructor. Learners’ ways of communication in online fora did not simply transfer from face-to-face to online environments, thus it is necessary to develop appropriate e-pedagogies (Li, 2008-2009). Tutty and Klein (2008), for example, suggested that face-to-face collaboration might be better suited for learning regarding well-defined facts and procedures. Virtual
collaboration might be better suited for learning regarding ill-structured problem solving. They also suggested use of asynchronous communication to large group and synchronous communication to small group. Veerman and Veldhuis-Diermanse (2001) reported that more off-task related messages occurred in synchronous fora and more task related, constructive, and high level of knowledge construction messages occurred in asynchronous fora.

Consequently, it is crucial to analyze the type and the quality of interaction and what factors influence discussion to promote the use of online fora in diverse situations. There are a number of variables and influences on the effectiveness of online discussion. The role of the instructor and the degree of the instructor’s intervention (Andresen, 2009; Coppola, Hiltz, & Rotter, 2002; Tutty & Klein, 2008), learner characteristics (Bullen, 1998), the nature of tasks (Dysthe, 2002; Puntambekar, 2006; Tutty & Klein, 2008), students’ participation (Dysthe, 2002; Moore & Marra, 2005), the structure of discussion and the discussion question (Bodzin & Park, 2000; Brooks & Jeong, 2006; Moore & Marra, 2005; Salmon, 2002), group composition (Guldberg & Pilkington, 2006; Tutty & Klein, 2008) and size (Chai & Tan, 2009), student perceptions (Puntambekar, 2006), for example were referred to as important influences on online discussion.

There have been many studies regarding asynchronous fora. However, online fora were used as a supplementary learning tool of small graduate level courses. Even though the majority (over 82 percent) of online courses had occurred at the undergraduate level (Allen & Seaman, 2010), there was little research focusing on undergraduate level (e.g., Bodzin & Park, 2000; Veerman & Veldhuis-Diermanse, 2001) compared with that of graduate level (e.g., Fahy, Crawford, & Ally, 2001; Moore & Marra, 2005; Pawan, Paulus, Yalcin, & Chang, 2003; Puntambekar, 2006). Most studies supported the educational significance of online fora focusing on sharing information, brainstorming or internalization of knowledge and presented practical implications for effective design of online discussion. However, these studies have indicated that most knowledge building activities were limited within lower phases.

There were also a few studies targeting large enrollment courses (e.g., Guzdial & Turns, 2000; Schellens & Valcke, 2005) because both instructors and students were concerned about both interaction and workload, and in these instances it was more difficult to engage students in online discussions than in face-to-face discussions. Guzdial and Turns (2000) measured the extent of over 1000 undergraduates’ participation in online environment and focused the participation dimension instead of knowledge construction. Schellens and Valcke (2005) confirmed that the interaction in a large online course with the weekly face-to-face session was very task-oriented and the amount of discussion activity was important influence on the nature and quality of the discussion and the phase of knowledge construction.

Further research is needed to enlarge the scope of its use in higher education regardless of the size of course and level of study. This study aimed to describe the major characteristics of interaction and participation patterns in a large online course. This study attempted to explain how the discussion question and evaluation criteria influenced the nature of discussion focusing on interaction and knowledge construction. The results will contribute to understanding online interaction and developing effective online instructional strategies.

**Methods**

**Case and participants**

This case study was conducted in an online undergraduate course in South Korea during the spring of 2008. This course was an elective course and it was entitled “Leadership Development”. This course lasted sixteen weeks and students learned diverse leadership theories. A total of 100 students enrolled in this course. 88 students actually participated in an online discussion and there were 38 females and 50 males. There were 7 freshmen, 14 sophomores, 26 juniors, and 41 seniors. There were 30 engineering, 24 arts and physical education, 19 natural science, 13 architecture, 1 humanities, and 1 social science major. The age of the students ranged from 19-25 years with a mean age of 22.3 years. Online discussion was scheduled for one week but it proceeded for two weeks (24 March-6 April) because of students’ low participation during the first week. Students could only communicate in an online environment and there was no face-to-face meeting. One female faculty member taught this course and one teaching assistant helped. The grade was decided by an individual task (case study for selected leader), examination, two online discussion sessions, and attendance. Each online discussion score was 5% of final grades. Online discussion was scheduled in week 4 and week 13.
This study focused on online fora in order to analyze students’ interaction; the first online discussion session was randomly selected. The discussion consisted of 10 groups of 10 students. Each group consisted of heterogeneous members considering gender, major, and grade. Each group should discuss which leadership theory was more effective: trait theory (Are leaders born?) or situational theory (Are leaders made?) considering their experiences, characters, and future jobs during one week. All 10 groups discussed the same questions on the bulletin boards at the same time but within their separate online group boards. The instructor posted names of group members and participation protocols or evaluation criteria before the online discussion started. Participation protocols or evaluation criteria specified how to participate in the online discussion, the deadline of discussion, and the minimum number and the type of posting. In other words, students should post their own thoughts and more than five responses to get a perfect score (5 point). Irrelevant messages did not get a score. The score was decided based on the number of initiations and response posting. Table 1 shows evaluation criteria.

<table>
<thead>
<tr>
<th>Number of Initiations</th>
<th>Number of Responses</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1-2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>3-4</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>More than 5</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1-2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3-4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>More than 5</td>
<td>5</td>
</tr>
</tbody>
</table>

Data collection

Online messages posted by students were collected as a main source of data. All students’ online discussions were automatically stored on the host computer because it was based on text-based communication. After the online discussion was finished, the researcher read postings several times and divided them into each group and discussion thread. The researcher interviewed four randomly chosen students and reviewed course evaluation to supplement the data. Interview contributed to understanding reasons for the patterns of interaction and participation and the meaning of online interaction. Table 2 shows backgrounds of four interviewees.

<table>
<thead>
<tr>
<th>Interviewee’s Name (Pseudonym)</th>
<th>Gender</th>
<th>Grade</th>
<th>Age</th>
<th>Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brian</td>
<td>M</td>
<td>Senior</td>
<td>23</td>
<td>Physical Education</td>
</tr>
<tr>
<td>Emily</td>
<td>F</td>
<td>Senior</td>
<td>22</td>
<td>Art</td>
</tr>
<tr>
<td>Kathy</td>
<td>F</td>
<td>Junior</td>
<td>21</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Eric</td>
<td>M</td>
<td>Sophomore</td>
<td>21</td>
<td>Civil Engineering</td>
</tr>
</tbody>
</table>

Data analysis

Each message was used as an individual unit of data analysis. Each individual message was coded according to eight variables (Author, Date, Group, Number of References, Number of Responses, Interaction Type, Interaction Function, and Phase of Knowledge Construction) for the examination of the patterns of interaction and participation, and the phase of knowledge construction. Eight variables were selected for explanation of the relationship between discussion question or evaluation criteria and knowledge construction. Author, date, group, and number of references automatically appeared in an online fora. Number of responses was calculated based on a message flow. All of the student names were replaced with pseudonyms.

Multidimensional analysis framework is necessary for in-depth understanding because online interaction is complex (Chai & Khine, 2006). Content analysis should be used to evaluate the quality of online discussion (Celentin, 2007; De Wever et al., 2006; Mazur, 2004). Many researchers (e.g., Gunawardena et al., 1997; Hara, Bonk, & Angeli, 2000; Henri, 1992; Newman, Webb, & Cochrane, 1995) have developed frameworks and indicators regarding content analysis. This study was based on Henri and Rigault’s (1996) content analysis for analysis of interaction type and function. Henri’s (1992) frequently used content analysis consisted of the following five dimensions. The first
was participative and was represented by the numbers of posted messages. The second was interactive and was divided into initiation and response. Initiation meant introduction of new topics or facilitation of discussion and came with response. Response included explicit response which directly responded to previous massage with calling name, and implicit response which expanded on a previous message. An independent message having no response was excluded in this study. The third was cognitive and meant exhibiting knowledge and skills related to the course content. It consisted of elementary clarification, in-depth clarification, inference, judgment, and strategies. The fourth was metacognitive and meant general knowledge and skills of the learning process and self-regulation of learning. The fifth was social and not related to the formal content of subject matter. Henri and Rigault (1996) added an organizational dimension which related to how to organize and accomplish a group project and course management. In this study, interaction type included initiation, explicit response, and implicit response. Interaction function consisted of cognitive, metacognitive, social, and organizational interaction.

Henri’s model was attractive because it helped to understand how interactivity including cognitive and metacognitive process of individuals occurs in a group (Lally, 2001). However, it was limited in explaining collaborative knowledge construction because it focused on quantitative aspects of participation (De Wever et al., 2006). Thus, this study was based on Gunawardena et al.’s five-phase interaction analysis model (IAM) to indicate the degree of the phase of knowledge construction activities. These successive five phases consisted of (1) sharing and comparing of information, (2) discovery and exploration of dissonance or inconsistency among ideas, concepts, or statements, (3) negotiation of meaning and/or co-construction of knowledge, (4) testing and modification of proposed synthesis or co-construction, and (5) agreement statements and applications of newly-constructed meaning. It was suited for describing a discussion flow and knowledge construction (Marra et al., 2004). The model has been used in many studies (e.g., Chai & Tan, 2009; Fahy et al., 2001; Moore & Marra, 2005; Schellens & Valcke, 2006) among online discussion forum transcripts analysis.

There were several different activities from indicators of each phase of IAM. The researcher modified indicators of each phase of knowledge construction (see Table 3) to categorize the data considering the meaning of each phase. The researcher and another rater coded all messages for data analysis to validate the coding procedures and the coders discussed and negotiated regarding the difference in coding to achieve agreement. Inter-rater reliabilities were calculated to express consistency. Cohen’s $k$ reliabilities for interaction type, function, and phase of knowledge building were .98, .84, and .69, respectively.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Stage</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Sharing/Comparing of Information</td>
<td>A</td>
<td>Citation of information/ Presenting opinion</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Agreement with other’s opinion</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Providing examples to support other’s opinion</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Asking detailed explanation regarding opinion</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Description of the discussion question</td>
</tr>
<tr>
<td>2: Discovery and exploration of dissonance/ inconsistency among ideas, concepts, or statements/supporting</td>
<td>A</td>
<td>Stating of disagreement</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Asking to clarify the reason of disagreement</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Restating own opinion or supporting by suggestion</td>
</tr>
<tr>
<td>3: Negotiation of meaning / co-construction of knowledge</td>
<td>A</td>
<td>Clarification of the meaning of terms</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>A statement of agreement and relative weight to disagreement</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Identification of specific disagreement</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Proposal of new statements embodying negotiation</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Proposal of integration</td>
</tr>
<tr>
<td>4: Testing and modification of proposed synthesis or co-construction</td>
<td>A</td>
<td>Testing against facts that participants already knew</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Testing against previous knowledge/concept</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Testing against previous experience or recent experience</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Testing against resource provided</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Testing against contradictory testimony in the textbook</td>
</tr>
<tr>
<td>5: Agreement statements/ applications of newly-constructed meaning</td>
<td>A</td>
<td>Convergence and summarization of participants agreements</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Application of new knowledge</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Statements by participants showing their understanding that they experienced critical reflection</td>
</tr>
</tbody>
</table>
Results and Discussion

Due date-centered participation

The total number of messages for the group discussion was 536 in a time span of 2 weeks. On average, each student wrote 3.36 messages per week. Compared with the findings of other studies, the participation rate is relatively high. Schellens and Valcke (2005), for example, found that 80 students posted 1428 messages during 12 weeks and individual student posted 1.48 messages per week. Chai and Khine (2006) reported that 11 teachers yielded 25.6 notes per week and each participant contributed about 2.33 notes in a week.

207 messages (38.6%) occurred in the first week and 329 messages (61.4%) were posted in the second week. There was an increase in posting over time and there was a sharp increase on the due-date (see Figure 1). The instructor notified an extension of period on March 28th and it might influence the decrease of number of daily postings on March 29th, the original due date. Due date, April 6th, was the highest for postings. This result showed the difference; there was a drop in posting over time in large enrollment course (Guldberg & Pilkington, 2007). Due date-centered participation pattern reflected Korean undergraduate students’ learning ‘style’ and the increase in posting over time meant the necessity of warming up.

![Figure 1. Number of Daily Individual Postings](image)

The asynchronous nature of discussion might force the students to procrastinate the discussion as well. The instructor notified students when they should participate in the discussion in advance of the discussion. However, they might stick to their own schedule irrespective of time frame as the quotes addressed and thus led to low participation during the first week, which resulted in an extension of period. This result confirmed that learners often forgot to participate in the discussion because of flexible time and place (Bullen, 1998). The result had implications regarding how long the instructor should give the students for online discussion. In other words, only providing ample time couldn’t facilitate online discussion considering students’ asynchronous participation. Setting an appropriate time frame needs further examination.

*I usually log in every Saturday and thus I was little upset when I saw only one message posted in our group board. We should post responses. Let’s do our best to encourage peers’ participation. I post my opinion first.*

*I didn’t know we should discuss because I didn’t check class syllabus. I think this is a limitation of an online course.*

Participation pattern based on individual time frame might lead to convergent participation where students posted their messages in a short time. During fourteen days, individual student’s participation was less than seven days and especially 1-3 days covered 79.6%. That is to say, many students posted the required number of messages at once and did not post any messages during the rest of time as table 4 shows. It might impact lack of sustainable discussion. This result suggested the evaluation focusing on quality, rather than quantity of interaction.
Table 4. Distribution of Participation by Day

<table>
<thead>
<tr>
<th>Day</th>
<th>Number of Participants</th>
<th>Percentage of Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>25.0%</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>20.5%</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>34.1%</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>10.2%</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>9.1%</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.1%</td>
</tr>
<tr>
<td>7-14</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

One-way interaction

There was a difference in participation rate according to the group. Group 6 was the highest of posting group with 68 postings. Group 1 was the lowest posting groups with 32 postings. The number of participants among groups was different according to each individual situation. There were 528 messages regarding interaction which covered 98.5% of the total messages. 441 messages (83.5%) were related to response and it looked like response-centered communication. However, most students responded to the peer’s first initiation within a topic and there was little turn-taking based on the message flow diagram. The diagram of message flow in most groups showed a similar pattern as in Figure 2. Each number meant group and individual participant.
In addition, all groups developed 82 discussion threads and 48 (58.5%) discussion threads belonged to monologues. This result confirmed serial monologues in which students freely represented their opinions and responded to previous messages with a minimal effort made to connect to the discussion threads rather than continuous inquiry presented by their peers covered most interaction in asynchronous discussion (Pawan et al., 2003). In other words, discussions in this study were one-way (replying to previous message) rather than two-way (replying to the reply). The density of the network based on message flow was quite low as well. There are several possible reasons for achieving sparse network. First, students did not have face-to-face contact and thus they did not have chance to build informal relationships. Second, only two weeks was not enough time to establish trust among students. Third, the number of group members was large and this might hinder developing virtual community.

Evaluation criteria might influence one-way interaction pattern. Students should post their own thought and more than five responses to get a perfect score and 64 students (74.4%) got 5 points. Accordingly, most students posted their perspectives as initiation and selectively responded to peers’ initiations without turn-taking to satisfy required number and type of posting. Actually, 1 initiation covered 76.1% and 5-6 responses covered 57.9% to satisfy evaluation criteria as Table 5 and Table 6 show.

<table>
<thead>
<tr>
<th>Table 5. Number of Initiations by Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Initiations</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6. Number of Responses by Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Responses</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Most students posted their message on each group forum and only two students posted their message on other group fora. Students had an opportunity to post their thoughts regardless of group but they did not actively participate in other group’s discussion. Students read most peers’ postings but posted a few responses to satisfy the required number of postings as Table 7 and Table 8 show. This result supported a previous study (Hara et al., 2000) which referred to type of posting of undergraduate students which was less responsive to their peers. Especially, a high percentage (66.0%) of messages having no response gave the implication regarding a ‘lurker’ who only read peers’ messages and did not present his or her own thoughts, and the degree of disconnected discussion. In this context, it was no wonder that the discussions were not adequately sustained and most interaction was one-way interaction as a result.

<table>
<thead>
<tr>
<th>Table 7: Number of References by Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of References</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>2~10</td>
</tr>
<tr>
<td>11~20</td>
</tr>
<tr>
<td>21~30</td>
</tr>
<tr>
<td>31~40</td>
</tr>
</tbody>
</table>

266
Actually, students selectively responded to peers’ posting according to their interest. They usually responded when they found a similar opinion and experience or had a different perspective. Students were also reluctant to engage in directly challenging other students’ opinions. It might result in a high percentage regarding a statement of agreement and relative weight to disagreement within phase 3. It might reflect the nature of the discussion question that asked students to select one leadership theory. Students always gave turn-taking to clarify their previous writing when peers’ asked for more clarification. However because of Confucian heritage students did not continuously criticize peer’s opinions after explanation and students did not want to be considered impolite in public. Openness of online fora based on text-based communication might discourage them from making controversial comments. This result might be related to the cultural norms of niceness in Asian countries considering Chai and Khine’s study (2006). Students often appreciated peer’s responses and asked for peer’s feedbacks considering selective response. It implied how students thought response in online interaction as the following quotes indicated.

Thank you for your response! I thought that my posting might be seemed aggressive if you focused only on the title. I wanted to facilitate discussion. Don’t feel sorry for me because this is exchanging of thoughts. I think that you responded to me without any bad feeling as I presented my opinion to peer without bad intention. I hope to share good opinions in the future.

First of all, I appreciate your reading and response. I want to explain what I wrote before. We learned several physical traits regarding trait theory according to the second week lecture note. I could misunderstand them. Please confirm and respond!

In general, I didn’t critically comment on members’ ideas. If somebody criticized each others’ opinions, he or she would look like a show-off or argumentative person.

### High portion of metacognitive interaction and higher phase of knowledge construction

Cognitive interaction covered 54.9 % of the total interaction related to the interaction function and showed online interaction was task-focused (Chai & Khine, 2006). Social and organizational interaction each covered 6.3% and 3.5% of the total interaction. Metacognitive interaction covered 35.3% of total interaction and high percentage of metacognitive interaction might be related to the nature of the discussion question. Discussion question required the selection and was confrontational as well. Students discussed which leadership theory is more effective between trait theory (Are leaders born?) and situational theory (Are leaders made?) considering their experiences, characters, and
future. Therefore, they had an opportunity for self-evaluation and connection theory with practice, and thus it might result in a high percentage of metacognitive interaction as the quote indicated.

Hello! My name is Bridget. I already posted several responses. Now I am posting the initiation. I was very shy and not sociable. I did not try without any help. My major is vocal music and I need a charisma for singing in public. I knew that sociability was important after I was a college student. I have faithfulness as a leader’s trait. I did not challenge the new task but I kept going if I started a task. I did not have confidence and determination. I think that trait theory is more effective than situational theory. Of course, it is important to learn in the process of life. However, the person who has a trait from a birth easier achieves good results.

Students sympathized and developed confidence to be a leader through sharing a similar experience. Sharing previous peers’ school experience and future plans sometimes cheered them up and changed their perspectives. Thus, students internalized what they already knew and newly learned through learner-learner interaction or had a chance to experience critical reflection as the following quotes illustrated.

I was focused on situational theory but I changed my mind. I don’t think that only one theory is the best. Some people are born with leader’s trait and others aren’t. I think that people should suitably consider both situational and trait theory according to the situation.

Your posting is good because it is based on your experience. Your experience becomes a valuable resource when you will be a leader. If you are used to the necessary task or relationship as a follower now, you will better understand and lead group members as a leader in the future. I agree with situational theory. I think if we make an effort to grow necessary attributes, we can be a good leader.

I am very confused between trait theory and situational theory. I preferred trait theory before. I think that 99% effort is valuable. I’m not sure which theory is better as time go on.

There were 493 messages regarding knowledge construction which covered 92.0% of the total messages. As Figure 3 illustrates, messages belonged to phase 1-2 covered 65% but a comparatively high percentage of messages belonged to phase 3. This study showed a result of 43%; 22%; 26%; 2%; 7% from phase 1 to phase 5 respectively. Comparing previous studies in which most messages were within Phase 1 and 2 (e.g., Chai & Tan, 2009; Fahy et al., 2001; Gunawaredena et al., 1997; Kanuka & Anderson, 1998), the result confirmed the potential of online fora in which the previous studies noted the messages belonged to phase 3 (Celentin, 2007; Moore & Marra, 2005; Schellens & Valcke, 2005). However, reviews of studies on addressing meaningful status of phase 3 stated that higher phase of knowledge construction are difficult to achieve. Schellens and Valcke’s study (2005), for example, reported a result.

Figure 3. Distribution of Each Phase of Knowledge Construction
of 52%; 14%; 33%; 1.2%; 0.4% from phase 1 to phase 5 respectively. Moore and Marra’s study (2005) also indicated a result of 56%; 22%; 19%; 3%; 0% from phase 1 to phase 5 respectively.

This study is slightly different than prior studies which most postings represented either phase 1 or 2 and very few postings in either section were coded in phase 4 or 5. Students focused on one issue within restricted time and they should choose to agree or disagree with controversial issue. Discussion question promoted students sharing their experience or challenging previous perspectives. It seems reasonable to assume that the discussion question and the structure of discussion might influence facilitating higher phase of knowledge construction.

Conclusions

The results could be summarized as follows: (1) Students posted required messages in a short time around the due-date and did not post any messages during other times within the restricted time. (2) Most interactions anchored the first initiation and had little turn-taking. (3) Students read many peer’s postings but selectively responded according to their interest. (4) There was a comparatively high portion of higher phase of knowledge construction and metacognitive interaction. (5) Discussion question and evaluation criteria influenced the pattern of interaction and participation, and knowledge construction.

The following instructional strategies could be suggested based on the results. First, this study showed a relatively high degree of interaction and participation even in a large enrollment course. Accordingly, this study provided instructors who teach in large online courses with confidence to attempt discussion in their course. Students more actively participated in discussion as time went on, and demonstrated in particular high participation around the due date. However, many students posted the required number of messages in a short time and spent most time reading and thinking messages without posting any of their own during online discussion. Thus it is difficult to state that high participation rate guarantees sustainable discussion. The result suggests that the density of network in term of students building on each other’s messages should be considered for successful asynchronous fora. Therefore, the instructor could assign students diverse roles such as summarizer, initiator, or opponent regarding encouraging their participation and prevent lurking.

Second, this study showed that students over-relied on their first initiation and did not develop discussion threads. Students posted their initiation regarding their own perspectives and selectively responded to peers’ opinions because of evaluation criteria. The result confirmed that evaluation criteria influenced on the pattern of interaction and participation. It is difficult to expect high level of knowledge construction within the structure of a forum having one first level note followed by two or three responses (Chai & Khine, 2006). The result showed that it was possible to obtain higher phase of knowledge construction within the structure of fora having serial monologue. However, this study recommended two-way interaction for achieving sustainable discussion and promoting higher phase of knowledge construction. Therefore, the instructor should pay more attention to the quality of interaction rather than the quantity and design the structure of forum having one first level note followed by lots of responses to prevent serial monologue in online interaction. Mandatory participation might lead to a psychological burden and have some unintended side effects as a previous study indicated (Bullen, 1998). Thus, the instructor should make an effort to decrease the negative effect of mandatory participation regarding facilitating responses. Delivering well-defined instructor’s expectations, or learning goals in advance will be effective strategies to promote responses.

Third, time is an important factor for sustainable discourse to be achieved. Chai and Tan (2009) emphasized the importance of allocating ample time for in-depth reflection and building relationships. Asynchronous interaction might be a barrier in delivering important notification or connecting messages. Therefore, the instructor should allow sufficient time of at least more than one week to warm up the discussion. However, simply providing students with enough time was not a panacea for active student participation considering the high percentage of participation around the due date and convergent participation. Thus, the instructor should recommend regular participation to achieve more sustainable discussion. The instructor should focus how many times students participated in the discussion rather than how many messages they posted as well.

Fourth, this study showed that a high portion of higher phase of knowledge construction, metacognitive, and task-oriented interaction. The task which should support one perspective between two confrontational theories asked students to retrospect their own experience, evaluate disposition, and connect theories learned with practice. It is not
surprising that most interaction was task-oriented considering the number of students, the time of discussion, and online environment without any face-to-face contact. This study confirmed the possibility that higher phases of co-construction of knowledge are possible to achieve focusing on the discussion question. Consequently, the result could contribute to overcoming previous studies (Chai & Khine, 2006; Chai & Tan, 2009; Järvelä & Häkkinen, 2002; Kanuka & Anderson, 1998; Murphy, 2004) that showed it is difficult to achieve higher phase knowledge construction in online discussion and rethink the previous study that smaller groups had a higher phase of knowledge creation (Schellens & Valcke, 2006). The result supported that the structure of discussion, the discussion question, and evaluation criteria are important component for successful online discussion regardless of the size of course. This was particularly true and helpful if the question encouraged students to reflect on a personal experience and asked them to express an opinion on a specific issue (Guldberg & Pilkington, 2007). Sharing personal experience might develop mutual connectedness and broaden students’ previous knowledge. Thus, the instructor should specifically structure the question of discussion that is more concrete rather than asking students to simply present their own perspectives for and experience negotiation through interaction.

Fifth, students usually responded to peers’ messages when they felt sympathy or had a different perspective because the discussion question included pro or con. Students always responded when peers wanted to clarify or post critical comments. The result also indicated that students did not actively criticize members’ ideas in public and completed without continuous in-depth turn-taking because of the cultural norms of politeness and it is consistent with comparable study of Chai and Khine (2006). Thus, the instructor should make participation protocol to prevent students from being ‘too nice’ to each other and continuously push them to critique peers’ perspectives. Grading weights reflecting critical comments or developing discussion can be used. The use of pre-structured threads could be considered regarding facilitating critical discussion (Brooks & Jeong, 2006).

Finally, this study confirmed the important role of the instructor. The instructor was referred as one of the important factors in successful online learning. Teaching in an online course is inherently different from the traditional face-to-face course (Andresen, 2009). The instructor should play a role as a cheerleader or motivator (Andresen, 2009; Dysthe, 2002). The instructor should maintain a minimal degree of intervention because students relied on the instructor’s feedback and learner-learner interaction could be decreased with active instructor participation as a result (Chai & Khine, 2006; Guldberg & Pilkington, 2007). Accordingly, the instructor should keep a balance to promote online discussion and design the structure of discussion in advance.

References


A Fuzzy Logic-based Personalized Learning System for Supporting Adaptive English Learning

Tung-Cheng Hsieh, Tzone-I Wang*, Chien-Yuan Su and Ming-Che Lee

Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan // 1Department of Computer and Communication Engineering, Ming Chuan University, Taoyuan, Taiwan // n9896132@mail.ncku.edu.tw // wti535@mail.ncku.edu.tw // n9897108@mail.ncku.edu.tw // leemc@mail.mcu.edu.tw

*Corresponding author

ABSTRACT

As a nearly global language, English as a Foreign Language (EFL) programs are essential for people wishing to learn English. Researchers have noted that extensive reading is an effective way to improve a person's command of English. Choosing suitable articles in accordance with a learner's needs, interests and ability using an e-learning system requires precise learner profiles. This paper proposes a personalized English article recommending system, which uses accumulated learner profiles to choose appropriate English articles for a learner. It employs fuzzy inference mechanisms, memory cycle updates, learner preferences and analytic hierarchy process (AHP) to help learners improve their English ability in an extensive reading environment. By using fuzzy inferences and personal memory cycle updates, it is possible to find an article best suited for both a learner's ability and her/his need to review vocabulary. After reading an article, a test is immediately provided to enhance a learner's memory for the words newly learned in the article. The responses of tests can be used to explicitly update memory cycles of the newly-learned vocabulary. In addition, this paper proposes a methodology that also implicitly modifies memory cycles of words that were learned before. By intensively reading articles recommended through the proposed approach, learners comprehend new words quickly and review words that they knew implicitly as well, thereby efficiently improving their vocabulary volume. Analyses of learner achievements and questionnaires have confirmed that the adaptive learning method presented in this study not only enhances the English ability of learners but also helps maintaining their learning interest.

Keywords

Intelligent tutoring systems, English learning, Fuzzy inference, Analytic hierarchy process

Introduction

With advances in network technologies, geographic barriers are hardly a problem now for global communication. Languages, whether written or spoken, are the major tools for cyber communication, and English, with its wide popularity, has been recognized as a global language. For non-native English-speaking people, extensive reading is a common way to improve a person's command of English. English is even taken as a major course in primary schools in many countries where English as a Foreign Language (EFL) is taught, especially East Asia. One of the keys to success in English learning depends on a person's vocabulary volume. Improving the English vocabulary of a learner has thus become a popular research issue in countries where EFL is widely taught.

English teaching in Taiwan usually emphasizes the analysis and memorization of stems, prefixes, and suffixes of unknown words and uses vocabulary in accordance with a learner’s command of English to explain the meaning of these words. This practice enables a learner to remember the meaning of a word, but it may not allow the learner comprehend the word and be able to use the word in different circumstances. As such, the learner may forget it quickly. To build up one's vocabulary volume quickly and sustainably, the best method is not to remember the words by rote but rather to read extensively and often. These experts all suggest that a long-term habit of extensively reading articles that are appropriate for a learner’s English ability can greatly improve the vocabulary and command of a learner of English (Song, 2000; Xuan, 2002; Chen & Hsu; 2008a). However, this strategy may be difficult to implement for a learner with no extensive vocabulary because the learner may have problems either in choosing appropriate levels of articles in accordance with her/his needs and interests or in figuring out the meaning of unknown words using the semantics of familiar words in an article that is obtained. Dictionaries are always helpful; however, the need to continually look up unknown words, which once learned may be forgotten in a few days as per learning curve theory, might also easily discourage a learner. Several studies have also developed language tutoring systems in order to assist learners in learning language (Heift & Nicholson, 2001; Hsu, 2008; Ferreira & Atkinson, 2009) and Essalmi et al. (2010) have also proposed different personalization strategies for personalized e-learning.

systems by using 18 personalization parameters. With regard to an e-learning system, it is also hard to select appropriate articles for such learners if details on or precise profiles of learners are never established.

This paper proposes an approach that can help a learner build up his/her English vocabulary volume by intensive article reading, during which meanings of unknown words are understood in the context of articles, paired with immediate tests, through which learners can further understand the vocabulary in the articles as well as can enhance the suitability of article recommendations. Beginners in English learning usually know only a limited number of vocabulary and phrases; it would be quite hard for them to figure out meanings of unknown words in an article beyond their ability. The proposed approach uses modern item response theory (IRT) to recursively calculate the current vocabulary ability of a learner and harnesses fuzzy logic and memory cycle theories to choose suitable articles from a pool of graded articles for the learner. After an article is read, the vocabulary in it, which a learner is supposed to have comprehended, is used to generate a test in order to check the degree to which the learner understands the article by focusing on vocabulary and phrases. A prototype system was built, and several experiments were conducted to evaluate the efficiency of the proposed approach. Analysis of learner achievements and questionnaires confirm that the proposed approach not only enhances the English ability learners but also enhances their learning interest as an effect of adaptive learning.

Related Works

Item Response Theory

Item response theory (IRT) has flourished in the context of modern test theories since 1950s. Many item response models have been established using different number of parameters. The most common one is probably the Rasch Model, which was proposed in 1960 in the field of psychometrics to study misreading in oral tests. Due to its simplicity and flexibility, it is widely used also in other applied fields, usually to analyze data from assessments of a person’s abilities, attitudes, personality traits, and so on. This study uses the one-parameter Rasch Model (Baker, 1992).

Fuzzy Logic Theory

The concept of fuzzy logic appeared following the development of the concept of a fuzzy set by Zadeh in 1965. Fuzzy sets serve as a means of representing and manipulating data that are not precise but fuzzy (Carlsson, Fedrizzi, & Fuller, 2004; Roshandeh, Puan, & Joshani, 2009). Fuzzy logic is a form of multi-valued logic that allows intermediate values to be defined between conventional evaluations such as true/false, yes/no, high/low, and big/small. Notions such as rather tall or very fast can be formulated mathematically and processed by computers in order to apply a more human-like way of thinking in computer programming (Hajek, 2006; Zadeh, 1984). Fuzzy logic theory addresses reasoning in the context of real-life uncertainty. For example, the concept of “a middle-age person” may be expressed as an interval, e.g., between 35 years and 55 years old, without a natural boundary in classical set theory. However, human interpretation allows a gradual transition between the categories of “old” and “too old” (Yen & Langari, 1998; Zimmermann, 1991). Figure 1 shows an example of the difference between a classical set and a fuzzy set. If the X-axis represents the age of a person and the Y-axis represents the degree of membership in a set, a classical set has a sharp boundary because a person either completely belongs to the “a middle-age person” set or does not belong to the set at all. In contrast, a fuzzy set is defined by a membership function that maps objects in a domain of concern to their membership value in the set (Yen & Langari, 1998). Membership functions are classified into several types, such as a bell membership function, a triangular membership function, and a trapezoid membership function. In this research, trapezoid membership functions are used to derive suitable article difficulty levels for a learner.

Recently, fuzzy logic has been applied to fields, such as control theory, in order to address artificial intelligence. Bai and Chen (2008) used fuzzy inference on data from learners’ assessments in automatically generating concept maps that have been subsequently employed in many adapting learning systems. Applying fuzzy logic, Ullah and Harib (2008) designed an intelligent material selecting method to identify the optimal material for robotic components at the initial stage of system designs.
Memory Learning Strategies

There are three stages of memory-processing involved in collecting, recognizing and learning information from environment, namely, short-term sensory memory, short-term memory and long-term memory, as shown in Figure 2 (Schmidt, 1991).
Short-Term Sensory Memory is a temporary buffer that holds information from the environment for about 20-30 seconds, after which it is lost. Selective attention determines what information moves on to Short-Term Memory. Short-Term Memory is a working memory with very limited capacity and duration; it holds a small amount of information in mind. Some rehearsed information is moved to the long-term memory and can be kept in mind, while other information is forgotten. Long-Term Memory can store information for a long time. To make use of information in long-term memory, information is moved back to short-term memory by using a process called retrieved.

The German psychologist Dr. Hermann Ebbinghaus proposed a well-known finding known as the forgetting curve in 1885. The forgetting curve is an exponential curve describing how fast a learner tends to forget the knowledge he/she has learned (Ebbinghaus, 1885; Loftus, 1985). A person tends to halve his/her memory of newly-learned knowledge in a matter of days or weeks unless the learned materials are consciously reviewed, as shown in Figure 3. Thus, a learner can only remember what he/she has learned persistently by rehearsing it. How long the learner should rehearse materials is very important in the learning process. If the review cycle is too short, a learner will waste his/her time reviewing materials still remembered. In contrast, the learned knowledge will be forgotten if the review cycle is long. Ho (2005) proposed a methodology to automatically adjust the interval for review of a learner in order to memorize several concepts. Its implementation involved a microprocessor-based mobile device. The timing for review was adjusted according to the memory cycles of different words of a learner, and based on these timing intervals, words repeatedly appear to intensify memorization. Chen and Chung (2008b) also proposed a memory cycle-based approach, which used modified formulae from Ho’s study to adjust memory cycles according to a learner’s ability, vocabulary difficulty, and the results of vocabulary assessment tests.

System Architecture

The proposed English article recommendation system uses accumulated learner profiles to calculate the English ability of a learner and choose appropriate articles for the learner using a fuzzy inference mechanism. It uses both
memory cycle updates and analytic hierarchy process (AHP) to help a learner in an extensive reading environment. By intensive reading of suitable articles, a learner can maintain their interest in learning English and, at the same time, improve his/her English ability efficiently. The system architecture is shown in Figure 4.

System Description

The major modules for carrying out the computation procedures are the Article Preprocessing Mechanism, the Degree of Vocabulary Cognition Calculation, the Memory Cycle Calculation, the Fuzzy Inference Mechanism, the Review Value Calculation and the Article Features Calculation. The descriptions of them are given as follows:

The Article Preprocessing Mechanism adds a newly collected article to the article database and computes correlations with other articles already in the database. It also calculates the frequencies of the words in the article and includes these values, along with the vocabulary, into the GEPT vocabulary database. A learner interacts with the system via a User Interface. To newcomers, the system provides a self-assessed questionnaire to identify their interests in different topics to create learner profiles in the user portfolio database.

For every news article collected, the Article Features Calculation module has five generators for computing the five main feature values, i.e., the Average Difficulty of Vocabulary (ADV), the Average Length of Sentence (ALS), the Total Length (TLA), the Average Ability of Vocabulary of a learner (AAV), and the Article Correlation (AC). The five feature values are used by the Fuzzy Inference Mechanism to generate a document fitness degree for each article for each learner. The Review Value Calculation module computes the review value for each article for a given learner by counting the words that should be reviewed by that learner. The higher the value is, the greater number of words the article contains that should be reviewed by the learner. The Document Recommendation Agent calculates the recommendation score (RS) of every article for a given learner according to that learner’s interest, the document fitness degree, and the review value of the article. The system suggests the article with the highest recommendation score to a learner.

After the learner reads the article, the system picks from the article words that the learner is supposed to have comprehended based on his/her ability level and generates an immediate test, which serves to check whether the learner has understood the meaning of newly-encountered words. The test result feeds back via the Learning Feedback Agent to the Degree of Vocabulary Cognition Calculation module for computing and updating the vocabulary abilities in the learner's profile in the user portfolio database; at the same time, the Memory Cycle Calculation module re-evaluates the memory cycles of the learner for the learned vocabularies that appear in the content of the article and updates them in the user portfolio database. When a learner logs in again, a suitable article, which contains words with memory cycles that are due to be reviewed, will be recommended by the Article Recommendation Agent for the learner. The details of each process are discussed as follows.

User Interface

User registration

A learner’s English vocabulary size matters much in learning English (Mochizuki & Aizawa, 2000); in this study, the Degree of Vocabulary Cognition (DVC) is regarded as a learner’s English ability. Table 1 shows the DVC used in this research with its corresponding values. To calculate the English ability of a first-time user, the Computerized Adaptive Testing (CAT) approach is adopted. Ten questions are tested. The vocabulary for CAT is selected from elementary, intermediate, and high-intermediate levels of GEPT; 50 questions are generated for each level. A question contains one English word and four Chinese translations, among which one is correct. According to the result of the CAT and using a one-parameter IRT model and the maximum likelihood estimation method, the vocabulary volume of a learner is evaluated.

The proposed system recommends English news articles of various topics collected from Internet for intensive reading. To address learners' Degree of Article Preferences (DAP), collected news articles are classified into eight categories, including Politics, Society, Sports, Business, Arts, Technology, Health and Travel. The proposed system uses a learner-assessment interface to interact with a newcomer in order to evaluate the preferences of the learner.
with respect to articles by using a value of $DAP$ computed from the answers of an assessment questionnaire. A Likert scale is adapted to represent the degree of article preferences for individual learners such that 0.1 indicates “Not at all; Very much dislike”, 0.3 is “Not all”, 0.5 is “Moderately”, 0.7 is “Well” and 0.9 is “Very well”.

Table 1. The degree of vocabulary cognition with its corresponding value

<table>
<thead>
<tr>
<th>Ranges of Vocabulary Ability of a Learner ($\theta$)</th>
<th>Corresponding values ($DVC$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2.0 \leq \theta &lt; -1.2$</td>
<td>0</td>
</tr>
<tr>
<td>$-1.2 \leq \theta &lt; -0.4$</td>
<td>0.25</td>
</tr>
<tr>
<td>$-0.4 \leq \theta &lt; 0.4$</td>
<td>0.5</td>
</tr>
<tr>
<td>$0.4 \leq \theta &lt; 1.2$</td>
<td>0.75</td>
</tr>
<tr>
<td>$1.2 \leq \theta &lt; 2.0$</td>
<td>1</td>
</tr>
</tbody>
</table>

User Learning

In this study, a web portal of the fuzzy logic-based personalized extensive English reading environment is developed. It consists of a set of hyperlinks and a left-side panel for logging in that displays the latest information on the system. In addition, a right-side panel for choosing article topics and displaying recommended articles also appears, as shown in Figure 5. After reading an article, a learner is asked for feedback, which evaluates the degree to which the recommended English article as understood, as shown at the bottom of Figure 5. More details on learning feedback are discussed in the next section.

Figure 5. The recommended English article and the feedback
Feedback on Learning

The learning feedback process involves three entities, namely, the learning feedback agent, the degree of vocabulary cognition calculation and the memory cycle calculation. The learning feedback agent obtains learning feedback from a learner and sends it for degree of vocabulary cognition calculation, which computes an ability value representing the current English vocabulary ability of the learner according to the feedback received. The learning feedback agent also triggers the memory cycle calculation to re-compute and update the memory cycles of the learned English vocabulary for an individual learner.

Learning Feedback Scales

In this study, a five-point Likert scale is used to indicate the extent to which a learner understands a recommended article, where 0.1, 0.3, 0.5, 0.7, and 0.9 indicate “Not at All”, “Not All”, “Moderately”, “Well”, and “Very Well”, respectively. Learner feedback helps to compute these values and also triggers the degree of vocabulary cognition calculations and memory cycle calculations.

Degree of Vocabulary Cognition Calculations

To continue presenting appropriate English articles to a learner, the system continuously monitors the English vocabulary ability of the learner by using the learning feedback collected. The Degree of Vocabulary Cognition is updated by using formula 1.

\[
DVC_{i,k}^{\text{new}} = DVC_{i,k}^{\text{old}} + \frac{\text{feedback}_{i,d} \times C}{\sum_{d=1}^{WT_{i,w}} DVC_{i,w}}
\]

(1)

where \( DVC_{i,k}^{\text{new}} \) is the new Degree of Vocabulary Cognition of word \( k \) for learner \( i \) after reading an article \( d \), and \( DVC_{i,k}^{\text{old}} \) is the original DVC of word \( k \) for the same learner \( i \). \( \text{feedback}_{i,d} \) represents the learning feedback that learner \( i \) gives after reading an article \( d \), and \( C \) is a constant set to 20 (Saragi, Nation, & Meister, 1978).

Memory Cycle Calculation

The Memory Cycle Calculation computes and updates the memory cycles of the words that have been learned by a learner. In addition to using the formulae for adjusting memory cycles for vocabulary adapted from Chen and Chung (2008b), this research infers the relationships between words in an article and intelligently adjusts the memory cycles of those words that do not appear in immediate tests but are known to a learner. If a word appears in a recommended article, its memory cycle is updated using formula 2 derived from Chen and Chung (2008b).

\[
MC_{i+1,w} = MC_{i,w} + \frac{\sum_{d=1}^{WT_{i,w}} \text{feedback}_{i,d} \times DVC_{i,w} \times F_{i+1}}{WT_{i,w}}
\]

(2)

where \( MC_{i+1,w} \) stands for the new memory cycle of the English word \( w \) for learner \( i \), and \( MC_{i,w} \) stands for the previous memory cycle of the English word \( w \) for learner \( i \). The memory cycle of a newly-learned word is initialized to zero in this study. \( F_{i+1} \) represents the Fibonacci sequence, and \( WT_{i,w} \) is the number of recommended articles that contain word \( w \) on a particular day \( t \) for learner \( i \). \( DVC_{i,w} \) represents the Degree of Vocabulary Cognition of the English word \( w \) for learner \( i \).

Article Recommendation

This study also maintains calculations on article topic preferences in learner portfolios. For intensive reading, fuzzy logic-based inference is used to find articles of a suitable difficulty level for a given learner. For each learner, the
system also computes a review value for each article, which is the percentage of vocabulary in the article a learner should review. The Article Recommendation Agent then combines these three criteria to calculate the article suitability level for every article for each learner.

**Fuzzy Inference Mechanism**

The fuzzy inference mechanism for finding a suitable difficulty level of articles for a learner consists of four steps, including the Input, the Fuzzifier, the Inference and the Defuzzifier (Lee, Jian, & Huang, 2005; Zimmermann, 1991), as shown in Figure 6.

![Figure 6. The steps of fuzzy inference mechanism](image)

**The Input Step**

In order to decide which article from the article database is the most suitable for a learner, the Article Features Calculation module computes five feature values for every article, including the Average Difficulty of Vocabulary (ADV), the Average Length of Sentence (ALS), the Total Length of Article (TLA), the Average Ability of Vocabulary of the learner (AAV), and the Article Correlation (AC). The ADV, ALS, TLA, AAV and AC values all relate to the content of an article, while AAV and AC are also related to a learner’s English ability and learning portfolio. The first step is the formation of Input Linguistic Features, which involves computing the five feature values for articles and processing them in the next step of the algorithm.

**The Fuzzifier Step**

This step computes the degree of membership for the linguistic feature values, i.e., the ADV, ALS, TLA, AAV and AC of each article. This study uses the trapezoidal membership function for each linguistic term. Each fuzzy input variable has three linguistic terms, namely, Low, Median, and High, each of which has a membership function to represent its degree of membership. Table 2 summarizes the membership functions of the linguistic terms of the five fuzzy input variables as defined for this research, and Figure 7 depicts an example of a membership function with respect to the feature variable ADV. The variables $d_1, d_2, u_1$ and $u_2$ represent the parameters which are used to define a trapezoid membership function of the linguistic term (Yen & Langari, 1998). For example, the membership function of ADV_Median is defined by four parameters, namely, $d_1 = 0.25$, $d_2 = 0.625$, $u_1 = 0.4$ and $u_2 = 0.425$ as shown in Figure 7.
Table 2. Membership Functions of Linguistic Terms of Fuzzy Variables

<table>
<thead>
<tr>
<th>Fuzzy Variables</th>
<th>Linguistic Terms</th>
<th>d1</th>
<th>u1</th>
<th>u2</th>
<th>d2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADV</td>
<td>ADV Low</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>ADV Median</td>
<td>0.25</td>
<td>0.4</td>
<td>0.425</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td>ADV High</td>
<td>0.425</td>
<td>0.625</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>ALS</td>
<td>ALS Low</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>ALS Median</td>
<td>0.2</td>
<td>0.3</td>
<td>0.425</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td>ALS High</td>
<td>0.425</td>
<td>0.625</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>TLA</td>
<td>TLA Low</td>
<td>0</td>
<td>0</td>
<td>0.035</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>TLA Median</td>
<td>0.035</td>
<td>0.25</td>
<td>0.3</td>
<td>0.525</td>
</tr>
<tr>
<td></td>
<td>TLA High</td>
<td>0.3</td>
<td>0.525</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>AAV</td>
<td>AAV Low</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>AAV Median</td>
<td>0.1</td>
<td>0.5</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>AAV High</td>
<td>0.6</td>
<td>0.9</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>AC</td>
<td>AC Low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>AC Median</td>
<td>0</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>AC High</td>
<td>0.05</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

DFL

| DFL Very Low    | 0   | 0   | 0   | 0.15 |
| DFL Low         | 0   | 0.15| 0.25| 0.4 |
| DFL Median      | 0.25| 0.4 | 0.55| 0.7 |
| DFL High        | 0.55| 0.7 | 0.8 | 1.0 |
| DFL Very High   | 0.8 | 1.0 | 1.0 | 1.0 |

Figure 7. Membership functions of feature variable ADV

The Inference Step

The third step is the inference step, which contains two parts, namely, the AND operation and the OR operation. It employs totally $3^5$, or 243, rules based on combinations of the three linguistic terms and five fuzzy input variables. Two such rules are shown in the bottom of Figure 6 as an example. The output variable of each rule is defined as a fuzzy output variable DFL, i.e., the Document Fitness for Learner, which includes DFL_Very_Low, DFL_Low, DFL_Median, DFL_High, and DFL_Very_High as its five associated linguistic terms. Their degree of membership is defined and shown in the bottom five rows of Table 2.

The Defuzzifier Step

This final step involves defuzzification. This paper uses the discrete Center of Area (COA) computation method, as shown in formula 3. Note that $DFL^*$ represents the DFL values after defuzzification, and $L$ is the number of quantization levels such that the finer the level is, the more precise the result is and the heavier the computation is (Lee, Jian, & Huang, 2005). In formula 3, $y_i$ is the $i$-th quantization value, and $D$ is the number of linguistic terms of
DFL. \( \mu_d(y_i) \) is the degree of membership of \( y_i \) belonging to \( d \). The result of the computation states how suitable an article is for a specific learner based on a quantized value between 0 and 1. The larger the value is, the more suitable an article is for a learner.

\[
DFL^* = \frac{\sum_{i=1}^{L} \sum_{d=1}^{D} \mu_d(y_i) \times y_i}{\sum_{i=1}^{L} \sum_{d=1}^{D} \mu_d(y_i)}
\]

(3)

**Review Value Calculation**

When rating an article for a learner, the system also considers the words contained in the article that the learner has to review. An article has a higher review value for a learner if it contains more vocabulary that the learner is due to review. Formula 4 is used to calculate the review value for each of the articles not yet read by a learner.

\[
RV_{i,j} = \sum_{w=1}^{RW_i} \frac{\text{Over}_\text{Review}_{i,w}}{RW_i \times D}
\]

(4)

In formula 4, \( RV_{ij} \) is the review value of article \( j \) for learner \( i \); the higher the value is, the more urgent learner \( i \) needs to review article \( j \). The value of \( RV_{ij} \) is normalized between 0 and 1. A review list recording the first \( n \) shortest memory cycle words, which the learner is due to review, is kept in a learner’s profile. \( RW_i \) is the weighting value of article \( j \) based on the review list such that the more words that need review appear in article \( j \), the higher the weighting value is. \( \text{Over}_\text{Review}_{i,w} \) is a value that represents the review delay for word \( w \) by learner \( i \); for instance, after a learner first learned word \( w \), which the system expects to be reviewed after 5 days, the learner comes across it on the 8th day in article \( j \). In this case, \( \text{Over}_\text{Review}_{i,w} \) then equals to 3 (8-5=3). \( D \) is set as the constant 10, which is the average times a learner must encounter a word in order to definitively remember the word (Chen & Chung, 2008b; Saragi, Nation, & Meister, 1978).

**Article Recommendation Agent**

The recommendation of an article to a learner is based on the article’s suitability for the learner. The computation of an article’s suitability involves \( DFL^* \), \( DAP \) and \( RV \). The well-known Analytic Hierarchy Process (AHP) developed by Saaty (1977, 1980) is used to determine the weights of these three values for each learner in this study. AHP has been widely used in a variety of fields, such as travel planning (Huang & Bian, 2009), medical and healthcare (Liberatore & Nydick, 2008; Sloane, 2004), and business (Yang et al., 2007). In this study, the evaluation matrix has been set as shown in Table 3, where \( w_1 \) denotes the weight for \( DFL^* \), \( w_2 \) denotes the weight for \( DAP \), and \( w_3 \) denotes the weight for \( RV \). This study sets \( w_1 = 0.643 \), \( w_2 = 0.283 \) and \( w_3 = 0.074 \).

<table>
<thead>
<tr>
<th>( DFL^* ) ((w_1))</th>
<th>( DAP ) ((w_2))</th>
<th>( RV ) ((w_3))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1/3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1/7</td>
<td>1/5</td>
<td>1</td>
</tr>
</tbody>
</table>

Formula 5 calculates the recommendation score (RS) of each article for a specific learner and suggests the article with the highest score to the learner.

\[
RS_{i,j} = w_1 \times DFL_{i,j}^* + w_2 \times DAP_{i,j} + w_3 \times RV_{i,j}
\]

(5)

**Experiment Results and Discussion**

Experiments have been conducted to evaluate the effectiveness of the method proposed in this study. Because not all of the sophomore students own computers, which will be more convenient for the web based system, we deliberated
and invited 30 of them as participants per group to verify the effectiveness of the proposed system in this study according to Airasian and Gay (2003). A total of 60 sophomore students, including 32 females and 28 males, from different departments of the National Cheng Kung University (NCKU), Taiwan were involved in experiments for four months.

**Experiment design**

A randomized pretest–posttest control group design was employed in this study. Figure 8 shows the experiment procedure. All participants were randomly assigned to either the experimental group (N=30) or the control group (N=30). The experiments mainly measure the effectiveness of the adaptive articles recommendations. After randomized grouping and pretesting, participants were engaged in learning processes provided by the developed system for four months. The performance of learners in the experimental group, who read articles recommended through the proposed method, was measured. The performance of learners in the control group, who read preferred articles and subjects at their will, was also measured in the same way. To evaluate their English ability and analyze their degree of satisfaction after the experimental procedure, the participating students were asked to take a posttest as well as complete a questionnaire.

![Figure 8. The experiment procedure](image)

**Experiment results and discussion**

This section presents three evaluation results based on learning feedback from learners, the learning performance of learners, and questionnaire feedback and analysis.

![Figure 9. Average of the learning scores from the top-ten learning activities](image)

**Learning feedback from learners**

Figure 9 shows a comparison of average learning feedback from both the experimental group and the control group. Each feedback indicates the degree of understanding of a learner on an article after he/she finishes the reading. The top-ten learning activities of each student were collected and analyzed for both groups in the experiment. For the 30 students in the experimental group, the average learning score ranged from 0.5 to 0.7. This indicates that learners in
the experimental group can understand most of the articles recommended by the system according to their learning abilities and preferences. In contrast, the average learning score ranged from 0.2 to 0.7 for the 30 students in the control group. This indicates that these learners sometimes understand the content of articles they chose, but other times they do not.

**Learning performance of learners**

Table 4 shows the average scores on the pretest and posttest for the experimental group and the control group. Table 5 shows that independent t-tests show that there are no significant differences in the pretest results ($t = .245, p = .807 > .05$) between experimental group and control group. Table 6 (A) and (B) compares the difference in the mean pretest and mean posttest for both the experimental group and the control group. The results indicate that there is no significant improvement in learners’ ability after the learning process for the control group ($t = -.083, p = .934 > .05$), while the improvement for the experimental group is significant ($t = -2.890, p = .007 < .05$).

<table>
<thead>
<tr>
<th>Table 4. The evaluation results of pretest and posttest for both the experimental group and control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Control group (N = 30)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Experimental group (N = 30)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. The t-test of the pretest of learning performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene's Test for Equality of Variances</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>F-test</td>
</tr>
<tr>
<td>Equal variances assumed</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6. The paired samples t-test of the pretest and post-test for experimental group and control group (A) Paired samples correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Control group (N = 30)</td>
</tr>
<tr>
<td>Experimental group (N = 30)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(B) Paired samples test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired differences</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Lower</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Control group (N = 30)</td>
</tr>
<tr>
<td>Experimental group (N = 30)</td>
</tr>
</tbody>
</table>
Questionnaire analysis

To further evaluate the usability of the system developed in this study, a questionnaire that contains 18 questions was used to survey the learners who participated in the experiments. It consisted of four parts entitled “Learning motivation and attitude”, “System operation”, “Degree of satisfaction in learning” and “Learner feedback”. Each question included five answer options, including ‘Strongly Agree’, ‘Agree’, ‘Neutral’, ‘Disagree’, and ‘Strongly disagree’. Appendix A presents further details on the questionnaires and evaluations. The questionnaire is highly internally consistent and has a 0.771 of Cronbach’s alpha coefficient, which indicates a higher reliability with a higher value. Among the 18 questions, questions No. 6, 7 and 8 are only included for the experimental group, and questions No. 9 and 10 are included only for the control group. For the control group, results reveal only 30% of the learners spend most of their time searching for suitable articles according to their English ability (Median = 2.97 < 3), while 43.3% of learners spend most of their time finding articles related to their preferences (Median = 3.23 > 3). This fact indicates that some of the control group learners spent most of their time searching articles they preferred to read rather than articles they could probably have easily understood. In contrast, the results show that most of the experimental group learners said that the proposed system not only offers appropriate English articles according to their abilities and preferences (40% and 43.3%) but also reduces their effort in searching articles they prefer (60%). Although this percentage is not greater than 50%, Table 7 shows that the medians for questions No. 6 and 7 are both more than question No. 3 (3.37 and 3.43).

<table>
<thead>
<tr>
<th>Question (#)</th>
<th>Samples</th>
<th>Total</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6</td>
<td>30</td>
<td>101</td>
<td>3.37</td>
<td>.850</td>
</tr>
<tr>
<td>Q7</td>
<td>30</td>
<td>102</td>
<td>3.43</td>
<td>.855</td>
</tr>
<tr>
<td>Q8</td>
<td>30</td>
<td>112</td>
<td>3.73</td>
<td>.785</td>
</tr>
<tr>
<td>Q9</td>
<td>30</td>
<td>89</td>
<td>2.97</td>
<td>.964</td>
</tr>
<tr>
<td>Q10</td>
<td>30</td>
<td>97</td>
<td>3.23</td>
<td>.858</td>
</tr>
</tbody>
</table>

Figure 10(A). Evaluation result of learner’s feedbacks of the question 17

Figure 10(B). Evaluation result of learner’s feedbacks of the question 18
Finally, Figure 10(A) and Figure 10(B) show the results of the survey regarding the advantages and disadvantages of the proposed system. Most of the learners (62%) thought that the proposed system could improve their English vocabulary ability. In contrast, 52% of participants suggested that there should be some listening and reading tests after they finish reading an article. As such, this provides a potential topic for future study.

Conclusion

In Taiwan, there are ongoing discussions and suggestions as to how to improve a student’s English ability. This paper proposes a unique approach that draws on the intensive reading to harness both fuzzy logic and memory cycle-adjusting methods in order to improve personalized English learning programs. The approach uses a questionnaire to understand a learner’s preferences and then uses fuzzy inference to find article of suitable difficulty levels for the learner. It then employs review values to compute the percentage of article vocabulary that the learner should review. It combines these three parameters to establish the article’s suitability formulae to compute the suitable level of articles for the learner. The system also uses memory cycle updating to adjust the memory cycles for words that a learner learns for the first time in a given article as well as the words that appear that need to be reviewed based on the learner’s learning feedback. The results of these experiments confirm that with intensive reading of articles as recommended by the approach, a learner can remember both new words and previously-learned words easily and for longer time, thereby efficiently improving the vocabulary ability of the learner.

Acknowledgements

This research was supported by the National Science Council in Taiwan under project NSC 99-2511-S-006-003-MY3.

References


Xuan, Y. Y. (2002). 1000 Words a Day. Classic Communication Corporate, Taiwan.


# Question Description

### Answers

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part A – Learning motivation and attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I believe I can improve my English ability by extensive reading.</td>
<td>18 (30%)</td>
<td>25 (41.7%)</td>
<td>14 (23.3%)</td>
<td>3 (5%)</td>
</tr>
<tr>
<td>2. I agree that using this system promotes my interest in reading English articles.</td>
<td>4 (6.7%)</td>
<td>19 (31.7%)</td>
<td>25 (41.7%)</td>
<td>11 (18.3%)</td>
</tr>
<tr>
<td>3. I believe using this system can improve my English reading ability.</td>
<td>8 (13.3%)</td>
<td>32 (53.4%)</td>
<td>13 (21.6%)</td>
<td>7 (11.7%)</td>
</tr>
<tr>
<td>4. I believe using this system continuously will effectively improve my English reading ability and speed.</td>
<td>8 (13.3%)</td>
<td>26 (43.3%)</td>
<td>19 (31.7%)</td>
<td>7 (11.7%)</td>
</tr>
<tr>
<td>5. I cannot easily comprehend the English article recommended by the system.</td>
<td>3 (5%)</td>
<td>22 (36.7%)</td>
<td>22 (36.7%)</td>
<td>13 (21.6%)</td>
</tr>
<tr>
<td>6. <em>I think that the system can offer me an appropriate suitable English article to read.</em></td>
<td>3 (10%)</td>
<td>9 (30%)</td>
<td>14 (46.7%)</td>
<td>4 (13.3%)</td>
</tr>
<tr>
<td>7. <em>I think that the recommended English article is in accordance with my interests.</em></td>
<td>3 (10%)</td>
<td>10 (33.3%)</td>
<td>13 (43.3%)</td>
<td>4 (13.3%)</td>
</tr>
<tr>
<td>8. <em>I believe this system greatly reduces the time I spend searching the Internet for articles that I prefer to read.</em></td>
<td>5 (16.7%)</td>
<td>13 (43.3%)</td>
<td>11 (36.7%)</td>
<td>1 (3.3%)</td>
</tr>
<tr>
<td>9. <strong>When using this system, I spend a lot of time searching for suitable English articles that are in accordance with my English ability.</strong></td>
<td>12 (40%)</td>
<td>7 (23.3%)</td>
<td>9 (30.0%)</td>
<td>2 (6.7%)</td>
</tr>
<tr>
<td>10. <strong>When using this system, I spend a lot of time searching for suitable English articles that are in accordance with my preference.</strong></td>
<td>3 (33.3%)</td>
<td>12 (40%)</td>
<td>7 (23.3%)</td>
<td>1 (3.3%)</td>
</tr>
</tbody>
</table>

* Experimental group only, ** Control group only

### Part B – System operation

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. I think the system provides a user-friendly interface.</td>
<td>10 (16.7%)</td>
<td>25 (41.7%)</td>
<td>20 (33.3%)</td>
<td>5 (8.3%)</td>
</tr>
<tr>
<td>12. I think the system has an easy-to-use user guide.</td>
<td>13 (21.6%)</td>
<td>25 (41.7%)</td>
<td>18 (30%)</td>
<td>4 (6.7%)</td>
</tr>
<tr>
<td>13. I completely understand how to use the system.</td>
<td>12 (20%)</td>
<td>25 (41.7%)</td>
<td>14 (23.3%)</td>
<td>9 (15%)</td>
</tr>
</tbody>
</table>

### Part C – Degree of satisfaction in learning

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>14. I will keep on using the system to help me learn English.</td>
<td>13 (21.6%)</td>
<td>23 (38.4%)</td>
<td>12 (20%)</td>
<td>9 (15%)</td>
</tr>
<tr>
<td>15. I am willing to introduce this system to my friends.</td>
<td>11 (18.2%)</td>
<td>23 (38.4%)</td>
<td>23 (38.4%)</td>
<td>3 (5%)</td>
</tr>
<tr>
<td>16. I believe that I can still improve my English ability by reading articles even if I do not use this system.</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>34 (56.8%)</td>
<td>22 (36.6%)</td>
</tr>
</tbody>
</table>

### Part D – Learner feedback

17. In your opinion, what are the advantages of this English article recommendation system?
   1. I think it can improve my English vocabulary ability.
   2. I think it can improve my English grammar ability.
   3. I think it can improve my English translation ability.
   4. Other

18. In your opinion, what are the disadvantages of this English article recommendation system?
   1. It would be better if the system could provide me with some other tests, such as listening test and reading test.
   2. It would be better if the system could provide me with a vocabulary/sentence translation function.
   3. It would be better if the system could provide me with a vocabulary/sentence pronunciation function.
   4. It would be better if the system could provide me with other articles. (Not just news articles)
   5. Other

---

288
Shared Mental Models on the Performance of e-Learning Content Development Teams

Il-Hyun Jo
Department of Educational Technology, Ewha Womans University, 11-1, Daehyun-dong, Seodaemun-gu, Seoul, 120-750 South Korea // ijo@ewha.ac.kr

ABSTRACT
The primary purpose of the study was to investigate team-based e-Learning content development projects from the perspective of the shared mental model (SMM) theory. The researcher conducted a study of 79 e-Learning content development teams in Korea to examine the relationship between taskwork and teamwork SMMs and the performance of the teams. Structural equation modeling (SEM) was used to analyze the parameter estimations. As hypothesized, the results indicated that interaction among e-Learning ID team members led to higher SMMs (Ed- this acronym has already been defined above) which in turn improved the team performance. Meanwhile, the interaction decreased with the progression of ID projects and with the role differentiation. The implications of the findings and directions for instructional design (ID) practices are discussed.

Keywords
Shared mental model, role division, team performance, e-learning

Introduction
In real world instructional design (ID) situations, team-based approaches are common. In e-Learning ID projects where a variety of expertise – e.g., ID, graphic design, and programming - is required, it is often difficult or impractical to find an all-in-one expert instructional designer. In this regard, Jo (2008) suggested that an e-Learning content development project is a typically team-based, ill-structured problem solving task that involves a series of complex, problem solving activities. However, ID settings as considered by most traditional ID theories and models are more logical or individual than collaborative or team-based (Jo, 2008). Most ID theorists, regardless of their epistemological perspectives, assume that their typical research targets are individual designers, not teams.

This discrepancy between the theories and real world practices may generate severe challenges to the ID research in e-Learning. Without the provision of relevant theories that explain the team aspects of ID practices, our credibility as application scientists might be challenged. There is growing evidence that the existence of shared mental models (SMMs) among the members of a work team has a great impact on team processes and task effectiveness (Klimoski & Mohammed, 1994; Mathieu, Heffner, Goodwin, Cannon-Bowers, & Salas, 2005). SMMs are socially constructed cognitive structures that represent shared knowledge or beliefs about an environment and its expected behavior (Klimoski & Mohammed, 1994). They influence team member behavior and improve coordination by enabling members to anticipate one another’s actions and needs (Cannon-Bowers, Salas, & Converse, 2005). This notion is particularly important when work events are unpredictable or when frequent communication is difficult (Mathieu et al., 2005), such as in the development of an e-Learning instructional design project (Jo, 2008). Some empirical studies have examined the relationship between SMMs and the team-based software design processes (e.g., Espinosa, Kraut, Lerch, Slaughter, Herbsleb, & Mockus, 2001). However, no reported research has investigated the effect of SMMs in team-based e-Learning ID projects.

The purposes of this study are; 1) to suggest a theoretical Model to explain the team-based ID processes in e-Learning content development project settings using the SMM perspective, and 2) to empirically validate the Model and investigate the path relationships among the Model’s structural factors. The results will provide theoretical and practical implications to the increasingly popular team-based ID practices.

Literature review and hypotheses
Shared Mental Model (SMM) and Team Work
Humans create representations of their worlds that are simpler than the entities they represent (Johnson-Laird, 1983) in order to reduce uncertainty in their lives (Klimoski & Mohammed, 1994). These representations, which are called
mental models, are cognitive structures that include specific types of knowledge humans use to describe, explain, and predict their surroundings (Rouse & Morris, 1986). Uncertainty is reduced through a heuristic function that individuals use to classify and retrieve the most salient pieces of information about situations, objects, and environments from their mental models (Cannon-Bowers et al., 1993). This process of identifying potential outcomes further reduces uncertainty. A collection of individuals working together as a team also needs mental representations, or SMMs, in order to effectively accomplish their assigned tasks.

Thus, the key for a team with diverse expertise to process information more effectively is to generate common understandings or SMMs. SMMs are "knowledge [and belief] structures held by members of a team that enable them to form accurate explanations and expectations for the task, and in turn, to coordinate their actions and adapt their behavior to demands of the task and other team members" (Cannon-Bowers et al., 1993; 228). These cognitive structures are expected to influence the way in which individuals cognitively process new information, both the content of what they process and the speed with which they are able to process it (Walsh, 1995). Thus, by shifting their focus from the individual level to the team level, team members can be better able to complete the project in a manner that is desirable for themselves, their teammates, and the organization.

Empirical studies that investigated the SMMs (e.g., Marks, Zaccaro, & Mathieu, 2000) suggest that team SMMs allow members to anticipate one another’s actions and coordinate their behaviors, especially when time and circumstances do not permit overt and lengthy communication and strategizing among team members. Teams who share mental models are expected to have common expectations of the task and team, allowing them to predict the behavior and resource needs of team members more accurately (Cannon-Bowers et al, 2005). Under the team and task circumstances such as e-Learning content development projects, members in teams must rely on preexisting knowledge to predict the actions of their team members and respond in a coordinated fashion to urgent and novel task demands in order to be more productive (Jo, 2008).

Multiple SMMs: Taskwork and Teamwork

The theoretical literature on SMMs suggests that the members of a team are likely to hold not one, but multiple SMMs (Klimoski & Mohammed, 1994). Although there are many detailed breakdowns of mental-model types, these models can be viewed as reflecting two major content domains: those about taskwork and those about teamwork (Cooke, Salas, Cannon-Bowers, & Stout, 2000; Klimoski and Mohamed, 1994).

Taskwork encompasses all activities related to the execution of the task, while teamwork encompasses all activities necessary for teammates to work with each other. Each of these may have different effects on coordination depending on the task. A taskwork SMM describes the content and structure of the team’s specific tasks. A teamwork SMM refers to how team members should interact with each other to accomplish the task and has been adopted by many researchers for representation because different types of projects have similar teamwork SMM content (e.g., Johnson & Lee, 2008). This division is also consistent with the idea that teams develop two tracks of behavior: a teamwork track and a taskwork track (McIntyre & Salas, 1995).

Research Hypotheses

As previous studies suggest, in project teams, members with different mental models about how tasks should be completed experience difficulty in coordinating their activities. To resolve these differences, team members need to exchange enough information in order to negotiate a mutually agreed upon solution and the means of achieving it. As information is accumulated through interactions such as observation, hearing others' explanations, or adapting one's own models, group mental models are thought to converge over time (Johnson-Laird, 1989; Klimoski & Mohammed, 1994; Mathieu et al., 2005).

Thus, researchers insist that interactions among members are a strong facilitator for the creation of the SMMs. The more team members communicate with each other, the more likely they are to form a common frame of reference and develop an SMM (Klimoski & Mohammed, 1994; Lurey & Raisinghani, 2001). Empirical research indicates that interactions among organizational members lead to similar interpretations of team- and task-events (e.g., Schein, 1992). In summary, relevant theories and empirical studies suggest that as team members develop experience with
the task and with other team members through communication and shared interest, they develop SMMs. These research findings lead the researcher to formulate the following two hypotheses:

Hypothesis 1. Member interactions will facilitate the development of the teamwork SMMs.
Hypothesis 2. Member interactions will facilitate the development of the taskwork SMMs.

As hypothesized, interaction seems to be a facilitator for the SMMs. However, the tradeoff is that the team member workload increases with increased interaction. Research indicates that SMM influences team performance by decreasing the communication demands, thereby allowing team members to allocate cognitive load to the task at hand (Lagan-Fox, Anglim, & Wilson, 2004). According to Donnellon and colleagues, the SMM evolves as the team undergoes a complex, iterative process only until they converge to a point that allows the team to function as a collective (Donnellon, Gray, & Bougon, 1986). Thus, once team members develop SMMs to a sufficient degree, there is little incentive to continue interactions that consume precious time and cognitive load that would be better used for more taskwork purposes. The professional e-Learning content development teams that have certain levels of taskwork- and teamwork- SMMs through interactions should not require much time for their design projects except in the early stages, when members need to understand the uniqueness of the newly assigned project tasks. Therefore, for professional e-Learning ID team members, the researcher formulated the following hypothesis:

Hypothesis 3. Project progress by month will negatively predict member interaction

Instances of reduced interaction and communication within groups may inhibit the exchange of task- or team-focused information, and thus delay or otherwise interfere with the creation of team-level cognition. Such a situation can emerge when members decide to work independently of one another and have little role overlap. In group situations, the task structure or degree of role differentiation is a critical factor affecting the amount of interaction (Reichers, 1987), because team members communicate differently based on how their roles are structured (Rentsch & Hall, 1994).

As noted by Edmondson (1999), the reflection and discussion required for team learning might also reduce team efficiency, a necessity in short-term project teams working to meet a deadline. In this regard, Druskat and Kayes (2000) report an interesting phenomenon found in MBA team project groups. In their study, the requirement for MBA students to meet deadlines and achieve high performance in project teams resulted in short-term performance goals taking precedence over interactions and learning (Druskat & Kayes, 2000). In another study of the effects of structure on team interaction, teams in which every member had the opportunity to perform all of the subtasks interacted significantly more than teams in which the responsibilities for the tasks were divided among the members (Urban, Bowers, Monday, & Morgan, 1995). Since group interaction to coordinate work is partly a function of the type of structure or division of labor within it, there may be situations that are less conducive to the formation of SMMs. Hence, the researcher formulated the following hypothesis:

Hypothesis 4. Role differentiation in teams will negatively predict member interaction.

As the preceding discussion implies, SMMs in a team should improve task performance, other conditions being equal. However, few researchers have examined the influence of the two types of SMMs, teamwork and taskwork, on team performance. In a laboratory study of two-person teams, Mathieu and colleagues assessed the team members’ SMMs and found that both taskwork and teamwork SMMs were significantly and positively related to team processes, which were in turn significantly related to team performance. However, the direct relationship between SMM and performance was not significant (Mathieu et al., 2000). In a similar, but more recent, laboratory-based study, Mathieu et al. (2005) showed that taskwork mental model similarity, but not teamwork mental model similarity, was significantly related to both team processes and team performance. Building on existing studies (e.g., Cannon-Bowers et al., 1993) and following the report of Mathieu et al. (2000, 2005), the researcher argues that both taskwork and teamwork SMMs enhance team performance. Thus, the researcher formulated the following hypotheses:

Hypothesis 5. Teamwork SMMs will positively predict the team performance
Hypothesis 6. Taskwork SMMs will positively predict the team performance
Based on the theoretical implications and empirical evidence, the researcher developed a theoretical Model to describe the causal relational structure of the variables previously discussed. The Model is depicted in Figure 1.

**Figure 1.** Research model and hypotheses

**Method**

**Sample and procedure**

The unit of analysis in the present study is the team, not the individual members. Seventy nine (79) e-Learning content development teams in Korea comprising 511 members participated in this study. Part-time employees or student interns were not included. The response rate of the survey was 85 % (523 out of 614). Twelve responses were not included in the analysis due to missing answers and/or obvious carelessness. The typical teams were composed of instructional designers, graphic designers, programmers, and system engineers. The average team size was 6.47 members with a range of 3 to 21. Of the respondents, 89% were full-time employees, 96% had college or advance degrees, 34% had education or educational technology degrees and 25% had computer science degrees.

The typical e-Learning ID projects in this study were corporate-oriented (vs. school-oriented) in terms of target audience, and were utilizing systematic approaches to ID in line with the recommendation from the Korean Ministry of Labor, which financially supports the ID projects by government policy.

The study used a single cross-sectional design (Fraenkel & Wallen, 2008; p.300) to investigate the changes in the observed variables with one-time data collection. Since a preliminary survey with the sample indicated that a typical e-Learning content development project takes about 3 months, the sample teams were categorized into three groups according to the month of the project progress: 32 teams were 0 to 1 month old, 37 were 1 to 2 months old, and 20 were in their third month. The levels of SMM, team performance, role differentiation, and member interaction were measured by the relevant instruments.

**Measures**

*Shared mental models (SMMs)*

To measure the SMMs of the participating teams, a translated version of the instrument developed by Levesque and colleagues (Levesque, Wilson, & Wholey, 2001) was used. The items ranged from assessments of the team's communication processes (‘Most of our team's communication is about technical issues.’), to evaluations of the climate (‘Voicing disagreement in this team is risky.’), and views of the team's structure (‘Lines of authority in this team are clear.’). Items were assessed on a 5-point Likert scale. In addition, a number of questions were posed to make specific assessments of the team's progress, such as ‘What percentage of your project do you feel is complete?’ The reliability and validity of the translated instrument was confirmed with an item internal consistency test using Cronbach’s alpha and confirmatory factor analysis using SPSS 15 and AMOS 7, respectively. Finally, 20 items, 10 for each of the teamwork and taskwork SMMs, were selected. Overall, post-hoc alpha and root mean square error of approximation (RMSEA) of the final instrument were .88 and .96, respectively.
Although SMMs have traditionally measured knowledge structures, it has been claimed that the construct should allow for the notion of evaluative belief structures (e.g., Johnson & Lee, 2008). The work on cognitive consensus can assist in this regard. Consensus is a different construct from consistency (Mohammed & Dumville, 2001). Measures of consistency are indices of reliability or the proportional consistency of variance among raters. Examples of consistency indices include the Pearson’s correlation coefficient $r$. A high interrater reliability measured by $r$ can be obtained if ratings by k judges are different but proportional. Specifically, consistency indices evaluate the similarity of rank orderings of judges’ ratings. Therefore, high interrater reliability can be obtained even when there is little manifest agreement between judges. For example, one rater may use values 97 through 100, while another uses 1 through 4. Thus, a correlational analysis of these ratings calculated by Pearson’s $r$ would reveal perfect consistency or similarity in the patterns of the ratings, whereas an index of agreement would reveal minimum consensus.

To assess the extent to which SMMs had developed in each team, the researcher used a measure of intra-team similarity, instead of Pearson’s $r$, as an overall index of the within-team consensus (Cooke et al., 2000) that evaluates the within-group agreement ($r_{WG}$) (James et al., 1984). $r_{WG}$ is the most frequently used measure of agreement or consensus (Webber et al., 2000), and is represented mathematically as:

$$r_{WG} = 1 - \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{\sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - \bar{x})^2}$$

where $\bar{x}_j$ is the within-group interrater reliability for a group of k judges on a single item, $\bar{x}$ is the observed variance on $x_j$, and $\bar{x}_u$ the variance on $x_j$ that would be expected if all judgments were due exclusively to random measurement error (James, Demaree, & Wolf, 1984). $r_{WG}$ controls for response biases, such as leniency and social desirability, that tend to inflate measures of group agreement (James et al., 2000).

**Member role differentiation**

Whereas the mental model measure looks at the level of perceptual agreement across a variety of variables, role differentiation measures the variance within group roles to determine the division of labor, i.e., how much the team members shared the duties of instructional analysis, storyboarding, media development, and organization of the team tasks. For instance, each team member’s own contribution and each team member’s contribution to the project role are measured using a 5-point scale. If a team's overall assessment is that every member made a 'moderate' or 'very small' contribution, the variance will be low, and the role differentiation will also be low (i.e., they shared the task among all members). If instead, a team has one member rated as contributing 'a lot', and another as contributing 'very little' to the same role, the division of labor in the group will be higher. As with the SMM measures, the $r_{WG}$ value was used as the indicator of member role differentiation in this study.

**Team interaction**

Each participant was asked to rate how much they had worked with other members of their team during the period since their project had commenced using a 5-point scale that ranged from 1 ('not at all') to 5 ('a lot') for two different modes of interactions: face-to-face and electronic interaction such as email or internet chatting. The team interaction score was calculated for each team by taking the mean of its members’ interaction scores.

**Team performance**

Board members of the e-Learning companies evaluated their content development teams, based on their weekly presentations, progress reports and the strategies they intended to use in the next period. Board members made their judgments by indicating their level of agreement with statements such as “The team is very likely to meet its instructional design quality objectives,” “The team has predicted the reactions of its clients to its design strategy,” and “Compared to other instructional design plans and storyboards that I have read, this one is ...” [5-point Likert scale, with endpoints of ‘unacceptable’ to ‘outstanding’]. The board evaluation score was calculated as the mean of multiple evaluation questions, averaged over all board members. The average reliability of this measure across all judges for a team was high for all three time-periods (alpha = .87, .89 and .91, respectively).
Data analysis

To test the model fit and the six individual hypotheses proposed, a structural equation modeling (SEM) analysis was conducted using AMOS 7. This analysis enabled the measurement error to be controlled for by fixing the random error variance to the product of the variance of the measured variable and the quantity one minus the estimated reliability for each variable. The quantity from the latent to the measured was fixed at one (1).

Before the hypotheses testing, preliminary analysis of the data revealed a few violations of normality measured by Shapiro-Wilk’s univariate normality, and AMOS’s multivariate normality index. In addition, the sample size was relatively small since the unit of analysis of the study was the team, not the individual. To deal with these issues, special care was required for the selection of parameter estimation method.

Thus, a maximum likelihood (ML) procedure with 2,000 iterations of bootstrapping option was used to estimate the model fit and other relevant parameter estimations. The rationale for using the ML procedure was two-fold: 1) ML estimation has been the most commonly used approach in SEM, and is therefore easily understood by the general readership, and 2) it has been found to be quite robust to a variety of less-than-optimal analytic conditions (e.g., small sample size, excessive kurtosis; Hoyle, 1995), as was observed in the present study. In addition, a bootstrapping procedure was selected as an option to overcome the study limitation caused by the relatively small sample size. Bootstrapping calculates the parameter estimates of interest resulting in an empirical sampling distribution. When the assumptions of the classical statistics such as a small sample size are severely violated, the empirical distribution that describes the actual distribution of the estimates made from this population will be substantially more accurate than the theoretical distribution.

Results

Descriptive statistics and correlation analysis

Descriptive statistics and correlation coefficients for all variables are reported in Table 1, which present a correlation matrix that permits the interested reader to recover the variance matrix. The reported data are rounded to three rather than the customary two decimal places to take full advantage of the precision offered by the SEM program, as recommended by Hoyle (Hoyle, 1995).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>1.971</td>
<td>.734</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role Differentiation</td>
<td>.932</td>
<td>.595</td>
<td>.140*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>6.232</td>
<td>3.891</td>
<td>-.070**</td>
<td>-.3.09**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team-related SMM</td>
<td>4.017</td>
<td>.539</td>
<td>-.012</td>
<td>-.052</td>
<td>.168**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Task-related SMM</td>
<td>4.654</td>
<td>.803</td>
<td>-.010</td>
<td>-.044</td>
<td>.144**</td>
<td>.024**</td>
<td>1</td>
</tr>
<tr>
<td>Team Performance</td>
<td>3.476</td>
<td>.741</td>
<td>-.011</td>
<td>.091</td>
<td>.114*</td>
<td>.767**</td>
<td>.203**</td>
</tr>
</tbody>
</table>

N=79 teams, *p<.05, **p<.01

Model fit

Major criteria fit indicators of the overall adequacy of the model fell within reasonable bounds. Table 2 summarizes the parameter estimates of the model fit accompanied by the major criteria indices. The fit criteria were referred from Hoyle (1995).

<table>
<thead>
<tr>
<th>NFI</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>.912</td>
<td>.921</td>
</tr>
<tr>
<td>Fit criteria</td>
<td>&gt;.900</td>
<td>&gt;.900</td>
</tr>
</tbody>
</table>
Individual hypotheses tests

Since the model fit was confirmed by the data, the individual hypotheses tests were allowed. All hypothesized relationships in the theoretical model were significant at the $p < .05$ level. The results are shown in Table 3.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Standardized estimates</th>
<th>CR</th>
<th>$p$</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1. Interactions $\rightarrow$ (+) Team SMMs.</td>
<td>.078</td>
<td>2.606</td>
<td>.009</td>
<td>Yes</td>
</tr>
<tr>
<td>H2. Interactions $\rightarrow$ (+) Task SMMs.</td>
<td>.049</td>
<td>2.542</td>
<td>.011</td>
<td>Yes</td>
</tr>
<tr>
<td>H3. Progress by the month $\rightarrow$ (-) Interactions</td>
<td>-1.126</td>
<td>-3.119</td>
<td>.002</td>
<td>Yes</td>
</tr>
<tr>
<td>H4. Role differentiation $\rightarrow$ (-) Interactions</td>
<td>-2.150</td>
<td>-2.493</td>
<td>.013</td>
<td>Yes</td>
</tr>
<tr>
<td>H5. Team SMMs $\rightarrow$ (+) Performance</td>
<td>1.317</td>
<td>5.888</td>
<td>.000</td>
<td>Yes</td>
</tr>
<tr>
<td>H6. Task SMMs $\rightarrow$ (+) Performance</td>
<td>1.551</td>
<td>7.510</td>
<td>.000</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Some Post-Hoc Observations

As reported previously, this study relied on data from self-reported surveys, which may be vulnerable to respondents’ subjectivity and social desirability. To triangulate the interpretation of the main data, a follow-up group interview with the team leaders of 18 of the 79 ID teams and 3 board members from three different companies was conducted. The interview data provided some additional information. First, according to the interviewees, the respondents had enough time to carefully read the survey instrument, which provided a relevant level of engagement when answering the survey. The relatively high response rate (85%) supports this notion. Second, the interviewees unanimously agreed that the most relevant evaluator of the performance of the ID teams was the board member. In Korea, e-Learning development companies are relatively small, so that the board members, directors and CEOs usually have opportunities to personally review and evaluate the design processes and the products.

Discussion and conclusion

This study’s major finding was that the Model developed on the basis of the SMM perspective was also relevant to a team-based e-Learning ID project’s setting: like other team-based projects, e-Learning content development projects can be better understood by employing SMM perspectives. Professional e-Learning ID project is a collaborative activity which requires human resources and coordination among functionally diverse members. Some team members may work on one task/aspect of an instructional or performance solution (developing assessment measures, for example) while other team members may work on a different ID problem (for example, storyboarding specific lessons). ID activities may be accomplished at different times as well as at different locations. In short, ID for e-Learning is a complex, collaborative enterprise, requiring shared mental models in order for goals to be achieved (Spector & Edmonds, 2002). Despite today’s advanced ID processes and development technologies, an ID team’s performance was strongly influenced by human-social factors such as shared mental models, or organized knowledge that members share about things like the task, and each other. This result suggests that SMMs and the social aspects of the ID should be considered in the research field of IDs. In addition to the general recognition of the criticality of SMMs for successful ID projects, three specific findings were highlighted by the study results. First, member interaction was a prime facilitator for the development of SMMs. Increased interaction among the members increased their SMMs. Traditionally, instructional designers have been interested in student interactions. However, the present study results suggest that our attention should expand to include the interactions in ID teams as well. Thus, interventions that evaluate and facilitate member interactions in ID teams should be developed. Evaluation methods such as a social network analysis (Jo, 2008) and facilitating interventions such as team development sessions (Salas, Rozell, Mullen, & Driskell, 1999) may be worth of study in the future.

Second, the SMMs developed through member interactions contributed to the team performance - the ultimate goal of the teams and the organizations. Even expert instructional designers and developers who are individually equipped with knowledge of ID models and with experience in diverse projects still need to develop SMMs when they start to work on new ID projects as a team. This demonstrates the important effect on team performance of applying extra efforts, individual and organizational, to the development of SMMs. Shared mental model training (Rouse & Morris, 1986), which aims to the development of members’ coordinated understanding of each other, shared goals and tasks,
may be considered an effective intervention at an early developmental stage of the ID teams. In the long run, however, more systematic and structural approaches that involve knowledge collection and sharing of knowledge activities (such as needs assessment, goal definition, prototype development, resource allocation, and so on) are required (Spector & Edmonds, 2002). As Jo (2009) reports, knowledge management system for ID professionals should be beneficial for the performance enhancement of ID teams.

Third, the degree of interactions in the teams decreased as members’ roles were differentiated by the division of labor. This finding is consistent with the reports of Druskat and Kayes (2000) and Urban et al (1995). For the Korean ID professionals participated in this study, the frequency of interactions with other members is negatively correlated with the level of role divisions. The present study findings suggest that such a short-term efficiency pursued by strict role division in teams has a negative influence on the amount of interactions that is critical for the knowledge sharing and reflection that occur in a team. This finding is consistent with research conducted by Langer (1997), which revealed that true collaboration is hindered by clearly mapped-out processes because team members reduce the level of mindfulness or the amount of thought and attention paid to the task processes. Thus, role division seems to be a double edged sword: it is necessary for highly standardized routines but not for ill-structured problem-solving tasks such as ID projects.

Fourth, the degree of interactions in these teams decreased as the projects progressed as well. This suggests that the time team members spent in meetings and other forms of communication decreased over the course of the project. For the Korean ID team members, interaction was instrumental for the development of SMMs but was cognitively costly as well. This finding is also consistent with Lagan-Fox et al (2004) and Donnellon et al (1986). Analyses of the present study left us with a number of implications that shared mental models help team members determine appropriate actions, form expectations of each other, explain how the team operates, describe the current state of the team, and predict its future state (Rouse et al., 1992). Organizations should attend to the shared, socially constructed causal connections that provide a blueprint for team action, and intervene early enough to mold these connections into effective shared mental models. The present analyses suggest that it is of particular importance that an e-Learning content development organization recognize the concrete and symbolic influence its culture and actions have on the ongoing development of shared mental models in teams.

The present research suffered a number of limitations. First, the study did not consider the ‘accuracy’ of the SMMs. Some theorists (e.g., Rentsch & Klimoski, 2001) have argued that shared and accurate SMMs enhance team performance. Second, the level of cross-validation was insufficient to strengthen the model validity. With a sample as small as that in the present study, even though the model fit indexes and p values met the set criteria, cross validation of the Model was necessary (Hoyle, 1995). Thirdly, politico-social mechanisms - power, authority, and norms – that could play a significant role in the collaboration processes were not included in this study. Fourthly, the study investigated the structure of the collaborations but not the regulation of the team interactions (e.g., Dillenbourg, 2002). Follow-up research will be required to answer practical question such as how the SMMs should be investigated. Lastly, due to the study context, the results may not be generalizable to non-Confucius cultures. Further studies are necessary with samples from other cultural contexts. Therefore, the study results must be interpreted cautiously.

References


Intelligent Discovery for Learning Objects Using Semantic Web Technologies

I-Ching Hsu
Department of Computer Science and Information Engineering, National Formosa University, Taiwan // hсутc@nfu.edu.tw

ABSTRACT
The concept of learning objects has been applied in the e-learning field to promote the accessibility, reusability, and interoperability of learning content. Learning Object Metadata (LOM) was developed to achieve these goals by describing learning objects in order to provide meaningful metadata. Unfortunately, the conventional LOM lacks the computer interpretability needed to support knowledge representation when searching for finding relevant learning objects. This study addresses this issue by defining a Multi-layered Semantic LOM Framework (MSLF) for integrating Semantic Web technologies into LOM. The proposed MSLF is used to develop LOFinder, an intelligent LOM shell that provides an alternative approach to enhancing the knowledge representations. To test its feasibility, this study implemented a java-based prototype of LOFinder that enables intelligent discovery of learning objects.

Keywords
LOM, Semantic Web, Ontology, RuleML, SCORM

Introduction
The Sharable Content Object Reference Model (SCORM) (ADL 2006) provides specifications for implementing e-learning systems and enabling learning object reusability and portability across diverse Learning Management Systems (LMS). The development and extension of SCORM metadata is based on IEEE Learning Object Metadata (LOM) (LOM 2005). The LOM in SCORM is used to provide consistent descriptions of SCORM-compliant learning objects, such as Content Aggregations, Activities, Sharable Content Objects, and Assets so that they can be identified, categorized, retrieved within and across systems in order to facilitate sharing and reuse.

The main problem with LOM is that it is an XML-based development, which emphasizes syntax and format rather than semantics and knowledge. Hence, even though LOM has the advantage of data transformations and digital libraries, it lacks the semantic metadata to provide reasoning and inference functions. These functions are necessary for the computer-interpretable descriptions, which are critical in the area of dynamic course decomposition, learning object mining, learning objects reusability and autoexec course generation (Kiu and Lee 2006; Balatsoukas, Morris et al. 2008). This is why most Web-based courses are still manually developed.

To improve the above problem, a mapping from LOM to statements in an RDF model has been defined (Nilsson, Palmer et al. 2003). Such a mapping allows LOM elements to be harvested as a resource of RDF statements. Additionally, RDF and related specifications are designed to make statements about the resource on the Web (that is, anything that has a URI), without the need to modify the resource itself. This enables document authors to annotate and encode the semantic relationships among resources on the Web. However, RDF alone does not provide common schema that helps to describe the resource classes and represent the types of relationships between resources. A specification with more facilities than those found in RDF to express semantics flexibly is needed. The Semantic Web (Shadbolt, Berners-Lee et al. 2006) can help solve these problems.

To enhance the knowledge representation of the XML-based markup language, the traditional Semantic Web approach is to upgrade the original XML-based to ontology-based markup language. The upgrade mentioned above from XML-based LOM to RDF-based LOM is an example. This approach is limited in that the original XML-based markup language has to be replaced with a new ontology-based markup language, causing the compatibility problems with existing data applications. This study proposes a novel integration approach that combines the first four layers of Semantic Web stack, namely URI layer, XML layer (LOM), ontology layer and rule layer. This integration approach is defined in a formal structure, called Multi-layered Semantic LOM Framework (MSLF), which is a specific sub-model of the Semantic Web stack for LOM applications. In MSLF, Semantic Web technologies can be integrated with LOM to enhance computational reasoning, and the original LOM can be retained to cooperate with ontologies and rules. That is, MSLF does not change the original schema of LOM. Hence, the existing LOM and SCORM metadata documents can continue to be used.
To demonstrate the feasibility of MSLF, an intelligent LOM shell for finding relevant learning objects, called LOFinder, is developed based on this framework. The core components of LOFinder include the LOM Base, Knowledge Base, Search Agent, and Inference Agent. It supports three different approaches for finding relevant learning objects of a certain course, namely LOM metadata, ontology-based reasoning and rule-based inference. Such dynamic finding is desirable for a number of reasons. Firstly, it is customized for each individual learning object, based on what metadata and knowledge the learning object has shown so far. Secondly, because the content or category of a learning object may keep changing, dynamic finding provides more up-to-date suggestions than a static design. Thirdly, as the number of learning objects may be large, adding suggestion links may become cumbersome for the course developer. Lastly, it can also be used at run-time to help in the decision of what content model component to deliver to the learner.

This study mainly aims to enhance the knowledge representation of LOM for computer-interpretable effects. It makes three main contributions. (1) This study defines a common framework MSLF for integrating Semantic Web technologies into LOM to facilitate machine understanding. (2) This study implements an intelligent LOFinder to demonstrate the feasibility of MSLF. LOFinder can be associated with various domain knowledge and metadata to offer dynamic relevant learning objects for different applications. (3) LOFinder can be easily transplanted to Learning Objects Repositories.

Semantic Web and Learning Objects

The section briefly discusses the current Semantic Web technologies and learning objects.

Semantic Web Stack

This study primarily focuses on the first four layers of Semantic Web Stack, including URI, XML, Ontology, and Rules layers. The first layer (including Unicode and URI) and second layer (including XML, Namespace, and XML Schema) represent current web technology. URI allows any web based resource to be identified. Unicode provides the basic character set for web based resources. XML and XML Schema provide a surface syntax for structured documents, but impose no semantic constraints on the meaning of these documents.

The third layer is ontology vocabulary that is rapidly becoming a reality through the development of ontology markup languages such as RDF, RDF Schema, DAML+OIL, and OWL (Web Ontology Language) (Smith, Welty et al. 2004). These ontology markup languages enable the creation of arbitrary domain ontologies that support the unambiguous description of web content. OWL is essentially an XML encoding of an expressive Description Logic, builds upon RDF and includes a substantial fragment of RDF-Schema. OWL has more facilities for expressing meaning and semantics than RDF and RDFS, thus OWL goes beyond these languages in its ability to represent machine-readable content on the Web. Unfortunately, these ontology markup languages are insufficient for describing the conditions under which specific relations might hold, which requires the explicit representation of implications, as is provided by logic programs, such as rules. A broad consensus has evolved in the Semantic Web community that the vision of the Semantic Web includes, specifically, rules as well as ontologies. The forth layer of W3C’s Semantic Web stack is rules to reflect this idea consensus view.

Learning Objects

The term learning object is one of the main research topics in the e-learning community in the recent years. The Semantic Web is an extension of the current web in which information is given well defined meaning, better enabling computers and people to work in cooperation. Many studies (Hsu 2009; Hsu, Chi et al. 2009; Hsu, Tzeng et al. 2009; Lu, Horng et al. 2010) adopt Semantic Web to build intelligent applications in various domains. Learning objects can be considered as resources that are accessible over the Internet. Therefore, Semantic Web can be used to enhance the accessibility, reusability, and interoperability of learning objects. In recent years, several research studies have focused on adopting ontology to enhance the interoperability of learning objects. But, they do not address the issue of how Semantic Web technologies can provide LOM to facilitate machine understanding. LOFinder is first to address the issue. The following gives a brief overview on existing ontology approaches.
(Gradinarova, Zhelezov et al. 2006) explain how Semantic Web technologies based on ontologies can improve different aspects of the management of E-Learning resources. Indeed, ontologies are a means of specifying the concepts and their relationships in a particular domain of interest. (Mohan and Brooks 2003) analyze relations of learning objects and the Semantic Web, especially emphasizing importance of ontologies. Accordingly, they identify several kinds of ontologies related to learning objects: ontologies covering domain concepts, ontologies for e-learning, ontologies about teaching and learning strategies, and ontologies about physical structuring of learning objects. (Wang, Tsai et al. 2007) propose an adaptive personalized recommendation model to help recommend SCORM-compliant learning objects. This model adopts an ontological approach to perform semantic discovery as well as both preference-based and correlation-based approaches to rank the degree of relevance of learning objects to a learner’s intension and preference. (Gasevic, Jovanovic et al. 2007) propose a framework for building learning object content using ontologies. In the previous work on using ontologies to describe learning objects, researchers employed ontologies exclusively for describing learning objects’ metadata. Although such an approach is useful for searching learning objects, it does not provide to reuse components of learning objects.

Multi-layered Semantic LOM Framework

Formal definitions of a Multi-layered Semantic LOM Framework (MSLF), which include Web Resource, learning object, LOM, Ontology, and Semantic Mapping Mechanism, are developed for general LOM applications. In particular, each layered component is formally described with respect to LOM.

**Definition 1. (Multi-layered Semantic LOM Framework)** This multiple-layered semantic framework for LOM applications can be formally defined, and is called the Multiple-layered Semantic LOM Framework (MSLF). The framework consists of a tuple $MSLF=<E, X, K, SMM, R>$ where

- $E$ denotes a collection of learning objects $LO$ in the URI layer (refer to Definition 2 for $LO$).
- $X$ denotes a collection of LOM-based metadata documents in the XML layer (refer to Definition 3 for LOM-based metadata document).
- $K$ denotes a collection of ontologies $O$ in the ontology layer (refer to Definition 4 for $O$).
- $SMM$ denotes the semantic mapping mechanism that defines how LOM can be combined with ontology-based knowledge (refer to Definition 5 for $SMM$).
- $R$ denotes a collection of rules $RLs$ in the rule layer (refer to Definition 6 for $RL$).

![Figure 1. Example of LOM application in multiple-layered semantic framework](image-url)
Figure 1 schematically depicts an illustrative example of program-markup language interactions in accordance with the framework, indicating how multiple-layered semantic Web technologies are adopted.

**Example 1.** Consider the concrete example given in Figure 1. Where $E$ is a set of learning objects and $X$ is a set of metadata coded with LOM. The software program ontology, $K$, is developed based on OWL and the logic program, RuleML (RuleML 2002), is taken for $R$.

**Definition 2. (Learning Object)** A learning object $LO \in E$ is a tuple $LO = \langle Rid, Uid, ALO \rangle$, where
- $Rid$ denotes an unique learning object identifier, such as URI.
- $Uid$ denotes an unique identifier of learning object owner.
- $ALO$ denotes a list of learning object attributes.

Learning objects are Web-accessible entities, such as web pages, pictures, programs, audio, video, etc. They are distributed in the Internet and are identified by URI.

**Definition 3. (LOM Document)** A LOM-based metadata document of a learning object is a tuple $LOMD = \langle LO, Ca, Re, Oth \rangle$, where
- $LO$ denotes a learning object.
- $Ca$ denotes a classification metadata that describes the meanings or abstract concepts of learning objects.
- $Re$ denotes a relation metadata that describes the meanings or abstract concepts of relationship between learning objects.
- $Oth$ denotes the other metadata in the LOM-based metadata document.

The LOM Information Model is broken up into nine categories. These categories are based on the definitions found in the LOM Information Model. The nine categories of metadata elements are: General category, Life Cycle category, Meta-metadata category, Technical category, Educational category, Rights category, Relation category, Annotation category, and Classification category. This study only focuses on Classification category and Relation category. The Classification category can be used to describe where the learning object falls within a particular classification system. The Relation category can be used to describe features that define the relationship between this learning object and other targeted learning objects. Each relationship is an outbound link that associates exactly two learning objects, one local and one remote, with an arc going from the former to the latter.

**Definition 4. (Ontology)** A web-based ontology $O \in K$ is a tuple $O = \langle C, P, \alpha, \beta, \gamma, \Sigma, \Pi \rangle$, where
- $C$ denotes a set of concepts representing classes in an ontology.
- $P$ denotes a set of relations representation properties in an ontology.
- $\alpha$ denotes the hierarchical relation function for classes. $\alpha : C \rightarrow C$, where
  - $\alpha(c_1) = c_2$ means that $c_1$ is a subclass of $c_2$.
  - This hierarchical relation can be used to determine if two classes have subclass/superclass relationship (Guarino and Welty 2004).
- $\beta$ denotes the hierarchical relation function for properties. $\beta : P \rightarrow P$, where
  - $\beta(p_1) = p_2$ means that $p_1$ is a sub-property of $p_2$.
- $\gamma$ denotes the attribute relation function between classes. $\gamma : P \rightarrow C \times C$, where
  - $\gamma(p_1) = (c_1, c_2)$ means that domain of $p_1$ is $c_1$ and range of $p_1$ is $c_2$.
- $\Sigma$ denotes a set of ontology axioms, expressed in an appropriate description logic.
- $\Pi$ denotes a set of RDF-based ontology language, such as RDF schema, DAML+OIL, or OWL.

An ontology is commonly defined as an explicit, formal specification of a shared conceptualization of a domain of interest (Studer, Benjamins et al. 1998). It describes some application-relevant part of the world in a machine understandable way. The reasoning capabilities of OWL will be discussed in the next section.

**Definition 5. (Semantic Mapping Mechanism)** A semantic mapping mechanism is a tuple $SMM = \langle LOMD, LLO, RLO, C, A, P, Y \rangle$, where
- $LOMD$ denotes a learning object metadata document (as defined in Definition 3).
- $LLO$ denotes a local learning object that is mainly described by the $LOMD$. 
• **RLO** denotes a remote learning object that is related to **LLO**.
• **C** denotes a set of concepts representing classes in an ontology.
• **Λ** denotes a classification mapping function. \( \Lambda: \text{LLO} \rightarrow \text{C} \)

The classification mapping function can make a classification tag (element) to refer an ontology class and acquire an additional semantic knowledge about the learning object.

\( \Lambda(\text{lo}_1) = c_1 \) means that the classification of the learning object \( \text{lo}_1 \) is set to class \( c_1 \).

• **P** denotes a set of properties in an ontology class.
• **Υ** denotes a relation mapping function. \( \Upsilon: \text{LLO} \times \text{RLO} \rightarrow \text{P} \)

The relation mapping function can make a relation tag (element) to refer an ontology property and acquire an additional semantic knowledge about the relationship.

\( \Upsilon(\text{lo}_1, \text{ro}_1) = p_1 \) means that starting learning object of \( p_1 \) is \( \text{lo}_1 \) and ending learning object of \( p_1 \) is \( \text{ro}_1 \).

In LOM, this study employs classification and relation metadata elements to provide extra semantic information. The classification element can indicate a character of the learning object has, while the relation element describes the meaning of the arc's ending learning object relative to its starting learning object. The content of both classification and relation elements are URI references that identify the learning object of the intended property. The format of this learning object is not standardized by LOM, and hence is open for proprietary semantic extensions. In the semantic LOM framework, each learning object can refer to an additional semantic through a classification element to set a specific ontology class. Similarly, each relationship between learning objects can refer to additional semantics through a relation element to set a specific ontology property.

**Example 2.** An example of instantiated LOM document is shown in Figure 1. Here **LLO** = cu-1, **RLO** = cu-4. Learning object references a specific ontology class through the classification mapping function as, \( \Lambda(\text{cu}-1) = \text{XML} \). In addition, the outbound arc defines a specific ontology property through the relation mapping function as in the following: \( \Upsilon(\text{cu}-1, \text{cu}-4) = \text{XMLParser} \).

**Definition 6. (Rule)** A rule \( RL \in R \) is a tuple \( RL = <C, P, H, B, Exp, RLP> \), where

- **H** denotes the head of the rule. \( H \subset C \cup P \)
- **B** denotes the body of the rule. \( B \subset C \cup P \)
- **Exp** denotes the rule of the form ‘if \( b_1 \) and \( b_2 \) and … and \( b_n \) then \( h \)’, where
  - \( h \in H \), \( b_1, b_2, … , b_n \in B \)
- **RLP** denotes a set of rule logic programs, such as RuleML, XRML (Lee and Sohn 2003), SRML(Thorpe and Ke 2003) or SWRL (Horrocks, Patel-Schneider et al. 2003).

The existing proposals for building a rule layer on top of the ontology layer of the Semantic Web refer to rule formalisms originating from the field of Logic Programming. The Rule Markup Language (RuleML) provides a natural mark-up for Datalog rules, using XML tags such as \(<\text{head}>\), \(<\text{body}>\), \(<\text{atom}>\), etc.

**Example 3.** In the example of Figure 1, the instantiated rule head and body can be depicted as \( h = \text{treeMode} \in P \), \( b_1 = \text{XMLParser} \in P \) and \( b_2 = \text{using} \in P \).

**Reasoning Capabilities of OWL and RuleML**

Before formally introducing LOFinder, the reasoning capabilities of OWL and RuleML are discussed. Both are essential components of LOFinder.

**Reasoning Capabilities of OWL**

The W3C OWL recommendation comprises three languages, given here in order of increasing expressive power: OWL Lite, OWL DL (Description Logic) and OWL Full. OWL Lite is a subset of OWL DL, which is in turn a subset of OWL Full, the most expressive language. OWL Full, extending RDF and RDF Schema to a full ontology language, provides the same set of constructors as OWL DL, but uses in an unconstrained manner. This study focuses on OWL DL, and adopts OWL to mean OWL DL without loss of generality. OWL is based on description
logic (Baader, Calvanese et al. 2003), which is a subset of first-order logic that provides sound and decidable reasoning support. The OWL allows the terminological hierarchy to be specified by a restricted set of first-order formulae. Additionally, OWL supports the development of an ontology multiple-layered architecture (Hsu and Kao 2005) that effectively integrates domain level ontologies with top level ontologies. These ontologies are a popular research topic in various communities. They provide a shared and common understanding of a domain that can be communicated between people and across application systems (Studer, Benjamins et al. 1998).

**XML-based RuleML Logic Program**

This knowledge is most appropriate represented using implication. With implication, the author can specify that Y (the consequence) is satisfied whenever X1...Xn (the antecedents) are all true. One of the simplest approaches for representing such implication is through rules expressions, which could be Horn clause statements. Unfortunately, general Horn clause statements are not explicitly representable using the primitives in OWL. OWL can represent simple implication as described in the previous section, but it has no mechanism for defining arbitrary, multi-element antecedents. For example, OWL’s description logics can not represent the following non-monotonic rule.

\[
\text{if XMLParser(XML,JAXP) and using(JAXP,DOM) then treeMode(XML,DOM)}
\]

Several researchers have shown how to interoperate, semantically and inferentially, between the leading semantic Web approaches using rules (for instance, RuleML logic programs) and ontologies (for instance, OWL description logics) by analyzing their expressive intersections (Grossof and Poon 2003).

**System Architecture of LOFinder**

LOFinder can be associated with various domain knowledge and metadata. The basic function of LOM is to provide metadata for e-learning applications. LOFinder supports three different approaches for finding dynamic correlation of learning object, namely LOM-based metadata, ontology-based reasoning and rule-based inference. The first approach, called LOM-based metadata, adopts XML-based LOM metadata to describe learning objects. This approach is used to develop existing e-learning applications, but still cannot intelligently locate relevant learning objects. The ontology-based reasoning approach provides OWL-based ontologies based on description logics to provide sound and decidable reasoning to LOM, and therefore can enhance the semantic reasoning capabilities of LOM. The rule-based inference approach can support inference capabilities that are complementary to those of an ontology-based reasoning approach. By building rules on top of ontologies, this approach can add intelligence based on logic programs.

The core components of LOM shell, LOFinder, include the LOM Base, Knowledge Base, Search Agent and Inference Agent. The flow-oriented LOFinder architecture is depicted in Figure 2.

- **LOM Base**: is an annotation repository composed of LOM-based metadata documents, which plays the same role as the fact base in a traditional expert system. A LOM-based metadata document is an XML document containing a set of relation metadata and classification metadata. The LOM Base is located on XML layer of the multi-layered semantic framework.

- **Knowledge Base**: is developed based on the Semantic Web technologies to support reasoning tasks, and plays the same role as the knowledge base in a traditional expert system. It is grouped into two categories: Ontology Base and Rule Base. The former corresponds to the ontology layer of the multi-layered semantic framework, while the latter corresponds to the rule layer. The Ontology Base comprises OWL-based ontologies for semantic reasoning. The Rule Base comprises RuleML-based rules to support a flexible and complex reasoning mechanism, which cannot easily be achieved by OWL-based ontologies.

- **Search Agent**: is a search engine that supports for a XPath query on an LOM-based metadata document collection within the LOM Base.

- **Inference Agent**: is an intelligent agent implemented based on a JESS-based rule engine (JESS 2007). It converts semantic of OWL-based ontologies and RuleML-based rules to JESS-based rules before starting an inference.
Figure 2. The flow-oriented LOFinder architecture

The information flow of the LOFinder occurs as follows.
1. The requester sends a learning object with URL to the search agent.
2. This step is the LOM-based metadata approach. The search agent relies on the request to query the LOM Base to finding all relevant LOM-based metadata documents of the learning object.
3. The transformation engine converts the LOM-based metadata to JESS-based fact.
4. This step is the ontology-based reasoning approach. The transformation engine conducts the following tasks to capture JESS-based rules:
   4.1 It retrieves and parses the relevant OWL-based ontologies quoted by the classification and relation tags.
   4.2 It converts these OWL-based ontologies to JESS-based rules.
5. This step is the rule-based inference approach. The transformation engine accomplish the following tasks to capture JESS-based rules:
   5.1 It relies on the relevant ontologies, mentioned on the step 4.2, to query the rule base to retrieve relevant RuleML-based rules.
   5.2 It converts these RuleML-based rules to JESS-based rules.
6. The JESS Rule Inference Engine derives new JESS-based facts from these existing JESS-based facts and JESS-based rules.
7. Finally, the inference agent passes the information of relevant learning objects, including LOM-based, ontology-based, and rule-based learning objects to the requester.

Domain Knowledge and Metadata Development

This section gives an example of a software domain to demonstrate how the multiple layers of semantic Web stack, i.e., LOM, OWL and RuleML, can be respectively mapped onto an LOM Base, Ontology Base, and Rule Base. The example in Figure 1 shows how this software domain is used in this study to provide the reasoning capabilities of LOFinder, which are further described in this section.

Ontology Development

The core ingredients of an OWL-based ontology include a set of concepts, a set of properties, and the relationships between the elements of these two sets. The present/introduce SoPro ontology offers the software language classification in a high abstraction level and is used to describe the semantic-based relation between classes, such as
markup language, program language, XML, XHTML, Java, JAXP, etc, involved in the software program domain. Figure 3 shows the semantic structure of SoPro ontology as a UML class diagram. The UML class diagram has as goal to give a graphical overview of the domain concepts and the relations among them. Every entity class and entity property in the diagram has already been described in detail.

The following four constraints present some partial code of the SoPro ontology to illustrate OWL-based description logics, including subclass, symmetric property, transitive property, and inverse property, respectively.

**Constraint 1. subclass**

OWL expression

```xml
<owl:Class rdf:ID="XML">
  <rdfs:subClassOf rdf:resource="#MarkupLanguage"/>
</owl:Class>
```

Semantic meaning

The XML is a subclass of the MarkupLanguage.

Rule expression

if XML(x) then MarkupLanguage(x)

**Figure 3. The UML Diagram for the SoPro ontology**

**Constraint 2. symmetric property**

OWL expression

```xml
<owl:SymmetricProperty rdf:ID="overlap">
  <rdfs:domain rdf:resource="#XML"/>
  <rdfs:range rdf:resource="#XML"/>
</owl:SymmetricProperty>
```

Semantic meaning

The overlap relation is symmetric.

Rule expression

if overlap(x, y) then overlap(y, x)

**Constraint 3. transitive property**

OWL expression

```xml
<owl:TransitiveProperty rdf:ID="include"/>
```

Semantic meaning
if x include with y, and y include with z then x include with z.
Rule expression
if include(x, y) and include(y, z) then include(x, z)

**Constraint 4. inverse property**

**OWL expression**
<owl:ObjectProperty rdf:ID="application"/>
<owl:inverseOf rdf:resource="#standard"/>  
<rdfs:domain rdf:resource="#XML"/>  
<rdfs:range rdf:resource="#XTHML"/>
</owl:ObjectProperty>

**Semantic meaning**
There is an inverse relation between application and standard.

**Rule expression**
if application (x, y) then standard(y, x)

**XML-based RuleML Logic Program**

The RuleML data model can represent Horn clause rules. An OWL class is treated as a unary predicate. An OWL property is treated as a binary predicate. Assertions about instances in a class are treated as rule atoms (e.g., facts) in which the class predicate appears. Assertions about property links between class instances are treated as rule atoms in which the property predicate appears. RuleML allows a predicate symbol to be a URI; this capability is used significantly herein, since the names of OWL classes are URIs.

Example 4 presents a RuleML rule for the non-monotonic rule that shows in the rule layer of Figure 1. This rule requires the representation of complex implications, a capability that goes beyond the simple implications available in OWL. This example indicates that RuleML not only provides a general implication in the form of Horn clauses, but also that its XML representation makes it the ideal choice for use with OWL. The SoPro ontology is defined in the previous section, and all its classes and properties are available to be used as elements in the rule. The RuleML rules file consists of only the LOFinder rulebase.

**Example 4.**

**RuleML expression**
<imp name="parser method">
  <_head><atom>
    <_opr><rel>treeMode</rel></_opr><var>XML</var><var>DOM</var>
  </atom></_head>
  <_body><and>
    <atom><_opr><rel>XMLParser</rel></_opr><var>XML</var><var>JAXP</var></atom>
    <atom><_opr><rel>using</rel></_opr><var>JAXP</var><var>DOM</var></atom>
  </and></_body></imp>

**Rule expression**
if XMLParser(XML, JAXP) and using(JAXP, DOM)
then treeMode(XML, DOM)

**Inference meaning**
If an XML course adopts a JAXP parser that is developed by DOM then the XML course adopts DOM method to extract data.

**LOM Metadata**

The software domain in the example represents all learning objects as HTML documents. The following discussion illustrates how LOM can be used to annotate relationships of learning objects using the concrete examples in the
In the URI layer, there are a number of learning objects, including XML Advance, XHTML Introduction, JAXP for XML, Java DOM, etc. In the XML layer, each learning object is described with a LOM-based metadata document that consists of classification metadata and relation metadata. These metadata can be summarized in Table 1 and Table 2 respectively. Each row in Table 1 denotes a classification concept in the LOM document, which is associated with a semantic clue that is the classification metadata of LOM. The value of a classification metadata can be mapped to an ontology class to inherit semantic knowledge. Each row in Table 2 denotes a relationship concept in the LOM document, which is associated with a semantic clue that is the relation metadata of LOM. The value of a relation metadata can be mapped to an ontology property to capture semantic relationships.

<table>
<thead>
<tr>
<th>learning object ID</th>
<th>classification: ontology#class</th>
<th>physical learning object</th>
</tr>
</thead>
<tbody>
<tr>
<td>cu-1</td>
<td>http://.../SoPro.owl#XML</td>
<td>http://.../xml.htm</td>
</tr>
<tr>
<td>cu-2</td>
<td>http://.../SoPro.owl#XHTML</td>
<td>http://.../xhtml.htm</td>
</tr>
<tr>
<td>cu-3</td>
<td>http://.../SoPro.owl#DOM</td>
<td>http://.../jdom.htm</td>
</tr>
<tr>
<td>cu-4</td>
<td>http://.../SoPro.owl#JAXP</td>
<td>http://.../jaxp.htm</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>start ID</th>
<th>end ID</th>
<th>relation: ontology#property</th>
</tr>
</thead>
<tbody>
<tr>
<td>cu-1</td>
<td>cu-4</td>
<td>http://.../SoPro.owl#XMLParser</td>
</tr>
<tr>
<td>cu-2</td>
<td>cu-1</td>
<td>http://.../SoPro.owl#standard</td>
</tr>
<tr>
<td>cu-4</td>
<td>cu-3</td>
<td>http://.../SoPro.owl#using</td>
</tr>
</tbody>
</table>

Two instances from the Table 1 and Table 2 are further revealed as follows. The classification metadata of "XML Advance" (i.e. cu-1) is annotated with the following metadata.

- learning object ID "cu-1" represents the "XML Advance" learning object.
- classification address the "XML Advance" is an instance of XML class. Therefore, "XML Advance" can inherit semantic knowledge from XML class.
- URL "http://sparc.nfu.edu.tw/~hsuic/el/xml.htm" provides physical location of "XML Advance".

![Figure 4. The query interface of LOFinder](image-url)
Similarly, the relation metadata of "XML Advance" is annotated with the following metadata.

- start ID "cu-1" is the starting learning object.
- end ID "cu-4" is the ending learning object.
- relation address the relationship (i.e. from cu-1 to cu-2) is an instance of XMLParser property. Therefore, the relationship can inherit semantic knowledge from XMLParser property.

**Relevant Learning Objects Discovery Using LOFinder**

This section will explicitly demonstrate the applicability of LOFinder by describing how the LOFinder can be adopted to provide LOM-based, ontology-based and rule-based relevant learning objects discovery using the software domain example as described in the previous section. In order to demonstrate the feasibility of LOFinder, a prototype java-based LOFinder is implemented in the paper. The user interface of LOFinder is shown in the Figure 4. A course creator selects a learning object and then presses the "Query" button. The LOFinder will rely on the learning object to invoke search agent and inference agent to retrieve the relevant learning objects information.

**LOM-Based Approach**

According to the learning object ID (i.e. cu-1) received, the search engine finds all relevant LOMs in the annotations base. Since all LOMs are actually XML documents, this corresponds to performing an XPath query on each LOM, looking for learning object whose identifier has the same value as "cu-1". The search result is shown in Figure 5. The cu-1’s LOM consists of an outbound link from cu-1 to cu-4. The LOM-based approach only depends on the cu-1’s LOM that exhibits a number of metadata. First, the file element describes the URL address of cu-1. Second, the classification element is used to describe where cu-1 acquires a particular ontology class. Third, the relation element is used to describe features that define the relationship between cu-1 and other learning objects. For example, the kind element describes where the relationship acquires a particular ontology property, and the resource element describes where cu-1 links to a particular learning object.

This kind of approach is that directly extracts data from the original LOM to retrieve the relevant information of learning objects, so this study calls it LOM-based metadata approach. The inference agent extracts data from the relation metadata of cu-1 to show that there is an XMLParser relation from cu-1 to cu-4. The output result of LOM-based approach is shown in (A) of Figure 4.

```
<?xml version="1.0" encoding="UTF-8"?>
<resource identifier="cu-1">
  <file href="http://sparc.nfu.edu.tw/~hsuic/el/xml.htm"/>
  <metadata><lom>
  ................
  <classification><purpose>
    <source>URI</source>
    <value>http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl/XML</value>
  </purpose></classification>
  <relation>
    <kind>
      <source>URI</source>
      <value>http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl#XMLParser</value>
    </kind>
  </resource><identifier>
  <catalog>learning object ID</catalog>
  <entry>cu-4</entry>
  </resource><identifier>
  <relation>
  .................
  </lom></metadata>
</resource>
```

*Figure 5. The partial LOM code of the learning object cu-1*
Ontology-Based Reasoning

The inference agent depends on semantics of SoPro ontology and cu-2’s LOM to reason the following facts.
1. The cu-2 learning object is an instance of XHTML class (see row 2 of Table 1).
2. There is a standard relation from cu-2 to cu-1 (see row 1 of Table 2).
3. The application property is an inverse of standard property.

Base on the above facts, inference agent can reason that there is an application relation from cu-1 to cu-2. The inference result is converted to an LOM document, as shown in Figure 6. The output result of ontology-based reasoning is shown in (B) of Figure 4.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<resource identifier="cu-1">
  <file href="http://sparc.nfu.edu.tw/~hsuic/el/xml.htm"/>
  <metadata>
    <lom>
      …………
      <classification><purpose>
        <source>URI</source>
        <value>http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl#XML</value>
      </purpose></classification>
      <relation>
        <kind>
          <source>URI</source>
          <value>http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl#application</value>
        </kind>
      </relation>
      …………
    </lom></metadata>
</resource>
```

Figure 6. The ontology-based metadata created by inference agent

Rule-based Inference

This inference agent relies on the previous inference results, LOM-based metadata documents and RuleML rules to perform the following tasks.
1. It converts the relation metadata of LOMs (see Table 2) and previous inference results (see Figure 6) to JESS-based facts, as shown in Figure 7.

```jess
(assert (triple (predicate "http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl#standard")
  (subject "cu-2") (object "cu-1")))
(assert (triple (predicate "http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl#application")
  (subject "cu-1") (object "cu-2")))
(assert (triple (predicate "http://sparc.nfu.edu.tw/~hsuic/sw/ontology/SoPro.owl#using")
  (subject "cu-1") (object "cu-4")))
```

Figure 7. The relation metadata are converted to JESS-based facts

2. It converts the RuleML rule (see Example 4) to JESS-based rule, as shown in Figure 8.
3. It relies on the above JESS-based facts and rules to infer the rule-based learning objects. The inference can infer that there is a treeMode relation from cu-1 to cu-3. The inference result is converted to an LOM document, as shown in Figure 9. The output result of rule-based reasoning is shown in (C) of Figure 4.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<resource identifier="cu-1">
  <file href="http://sparc.nfu.edu.tw/~hsuic/el/xml.htm"/>
  <metadata><lom>
    ……………………………
    ……………………………
  </lom></metadata></resource>
```

**Figure 8.** The RuleML rule is transformed to JESS-based rule

**Figure 9.** The rule-based metadata created by inference agent

**Experimental Results**

After describing the framework for enhancing the reasoning capabilities of LOM through LOFinder, a preliminary experiment is performed to test the expressiveness of the MSLF and the reasoning capabilities of LOFinder.

The test dataset contained 125 learning objects distributed in different classes of the SoPro ontology. In addition to the test dataset, there were 217 relations annotated in LOM-based metadata documents among those learning objects. Altogether, nine rules were identified as necessary to infer for the relevant learning objects. The complete list of rules can be found in Table 3. The first six rules are ontology-based reasoning, and the first three rules do not directly support to produce learning objects but can be referred by other rules. The last three rules are rule-based inference.

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Type</th>
<th>Rule expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule-1</td>
<td>ontology(subclass)</td>
<td>if XML(x) then MarkupLanguage(x)</td>
</tr>
<tr>
<td>Rule-2</td>
<td>ontology(subclass)</td>
<td>if XHTML(x) then MarkupLanguage(x)</td>
</tr>
<tr>
<td>Rule-3</td>
<td>ontology(subclass)</td>
<td>if HTML(x) then MarkupLanguage(x)</td>
</tr>
<tr>
<td>Rule-4</td>
<td>ontology(symmetric)</td>
<td>if overlap(x, y) then overlap(y, x)</td>
</tr>
<tr>
<td>Rule-5</td>
<td>ontology(transitive)</td>
<td>if include(x, y) and include(y, z) then include(x, z)</td>
</tr>
<tr>
<td>Rule-6</td>
<td>ontology(inverse)</td>
<td>if standard(x, y) then application(y, x)</td>
</tr>
<tr>
<td>Rule-7</td>
<td>rule</td>
<td>if XMLParser(x, y) and using(y, z) then treeMode(x, z)</td>
</tr>
<tr>
<td>Rule-8</td>
<td>rule</td>
<td>if XMLParser(x, y) and event(y, z) then eventMode(x, z)</td>
</tr>
<tr>
<td>Rule-9</td>
<td>rule</td>
<td>if template(x, y) and format(x, z) then style(y, z)</td>
</tr>
</tbody>
</table>
The experience conditions are described as follows:
1. Experiments were performed on a 2.4 GHz Pentium IV PC with 1024 Mb of RAM, running Linux.
2. All learning objects will be processed in random and one by one by LOFinder.
3. After a learning object has processed by LOFinder, the new inference facts could be kept in the memory and the running times could be added up for each rule that is triggered in the inference.

Total run time was 48.6 seconds. The search agent executed only one search of the LOM Base before extracting the relevant information needed to retrieve LOM-based learning objects, which was completed in only 1.1 seconds. Additionally, the remaining run time was spent inferring to get new learning objects, including ontology-based links and rule-based links. In total, 387 ontology-based links and 38 rule-based links were generated. The summary of test results is shown in Table 4.

<table>
<thead>
<tr>
<th>Link numbers</th>
<th>LOM-based links</th>
<th>ontology-based links</th>
<th>rule-based links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Times (ms)</td>
<td>1102</td>
<td>21235</td>
<td>26331</td>
</tr>
</tbody>
</table>

Figure 10 shows the execution time for each rule. The experimental results showed that the more complicated rule needs more running times. The inference time for rule 5 was increased by the addition of a transitive property. The inference agent must execute a complicated recursive function to derive the transitive result. Compared to unary predicates, the binary predicates such as rule4, rule5, rule 6, rule 7, rule 8, and rule 9, have longer inference times. The last three rules have longer inference times due to their numerous clauses and binary predicates.

**Summary and Concluding Remarks**

The LOM was developed based on the XML standard to facilitate the search, evaluation, sharing, and exchange of learning objects. The main weakness of LOM is its lack of semantic metadata needed for reasoning and inference functions. This study therefore developed LOFinder, an intelligent LOM shell based on Semantic Web technologies that enhance the semantics and knowledge representation of LOM. After introducing and defining the proposed multi-layered Semantic LOM Framework (MSLF) for LOM, the following discussion describes how the intelligence, modularity, and transparency of LOFinder enhance the discovery of learning objects.

Cloud computing is a newly emerging computing paradigm that facilitates the rapid building of a next-generation information systems via the Internet. One future work would be to extend LOFinder to support the intelligent e-learning applications in the cloud computing environment. Another future direction of development is upgrading LOFinder to a general framework for reusability, i.e., limiting the components of LOFinder to a LOM Base, Knowledge Base, Search Agent, and Inference Agent with no built-in domain knowledge in the Knowledge Base and with no domain metadata in the LOM Base. Furthermore, User-friendly interfaces are essential for enabling easy access to the LOM Base and Knowledge Base by domain experts.
References


A Model for Predicting Learning Flow and Achievement in Corporate e-Learning

Young Ju Joo, Kyu Yon Lim* and Su Mi Kim
Department of Educational Technology, Ewha Womans University, Seoul, Korea // youngju@ewha.ac.kr // klim@ewha.ac.kr // subi@ewha.ac.kr

*Corresponding author

ABSTRACT
The primary objective of this study was to investigate the determinants of learning flow and achievement in corporate online training. Self-efficacy, intrinsic value, and test anxiety were selected as learners’ motivational factors, while perceived usefulness and ease of use were also selected as learning environmental factors. Learning flow was considered as a mediator of predictors and achievement. Regarding methodological approach, structural equation modeling was employed in order to provide cause-and-effect inferences. The study participants were 248 learners who completed an e-learning courseware at a large Korean company and responded to online surveys. The findings suggested that self-efficacy, intrinsic value, and perceived usefulness and ease of use affected learning flow, while intrinsic value, test anxiety, and perceived usefulness and ease of use were significant predictors of achievement. The results revealed perceived usefulness and ease of use to be the most influential factor for both learning flow and achievement.

Keywords
Corporate e-learning, Self-efficacy, Intrinsic value, Technology acceptance, Learning flow

Introduction
E-learning has been around for more than a decade and has become widely regarded as a viable option for a variety of educational contexts. Especially, it forms the core of numerous business plans, as new technologies provide a new set of tools that can add value to traditional learning modes, such as accessibility to content, efficient management of courseware and learners, and enhanced delivery channels (Wild, Griggs, & Downing, 2002). In addition to these positive benefits, economic savings have made e-learning a high priority for many corporations (Strother, 2002). Given that as much as 40% of money spent on in-person corporate learning is eaten up by travel cost (Zhang & Nunamaker, 2003), companies using online training can expect plenty of time and cost savings, compared with conventional face-to-face training.

Despite the rapid growth of e-learning in the corporate training sector, this quantitative growth has not always guaranteed an equivalent improvement in the quality of learning. Especially, learners participating in online training in a corporate setting are likely to have their own job tasks to perform, which makes it difficult for them to concentrate on the learning itself. Hence, cognitive engagement of learners has drawn keen attention from researchers interested in the learners’ experience during online learning as well as the learning outcome (Herrington, Oliver, & Reeves, 2003). More specifically, learning flow has been reported as a construct related to learners’ engagement, predicting learning achievement. According to Csikszentmihalyi (1997), learning flow is characterized by complete absorption during learning. In other words, flow is the optimal experience as a mental state of extremely rewarding concentration that emerges in-between frustration and boredom. Flow becomes more important in the e-learning environment where there is no physical limitation in terms of time and space. When the learners do not experience flow, they may produce low engagement throughout learning, or even worse, fail to complete the e-learning (Skadberg & Kimmel, 2004). Considering that learning flow is a potential indicator of online learning achievement, more discussion on flow is necessary to expand our understanding of the phenomenon of corporate e-learning. The primary objective of this study is to investigate the determinants of learning flow and achievement in corporate online training.

Factors related to learning flow: Self-efficacy, intrinsic value, perceived usefulness and ease of use

Based on an extensive review of prior research, self-efficacy, intrinsic value, and perceived usefulness and ease of use of the e-learning program have been identified as critical variables predicting learning flow. Self-efficacy is a belief in one’s capabilities to organize and execute the courses of action (Bandura, 1977). Since these beliefs are...
determinants of how people think, behave, and feel, they play important roles during the course of learning. Zimmerman and Schunk (1989) described self-efficacy as an important factor that resides within the learner, mediates between cognition and affect, and affects academic performance. The relationship between self-efficacy and learning flow has been reported by previous studies. Meece, Blumenfeld and Hoyle (1988) examined the levels of self-efficacy of 275 fifth and sixth graders in a traditional classroom environment, and divided them into two groups of high and low self-efficacy. The former students showed higher outcome expectation, deeper engagement during learning for a longer period of time, and higher participation. In an online learning environment, Puzziferro (2008) investigated 815 undergraduate-level students’ self-efficacy and self-regulated learning skills, and reported that self-efficacy was a significant predictor of learning flow and achievement as well.

Learners’ intrinsic value has been identified as another important factor influencing learning flow. Intrinsic value is defined as the enjoyment one gains from doing the task (Wigfield & Cambria, 2010). Intrinsic value has been conceptualized in various ways (e.g., learning vs. performance goals, intrinsic vs. extrinsic orientation, task value, and intrinsic interest), but it essentially refers the reason for doing a task (Pintrich & DeGroot, 1990). When a learner is intrinsically motivated, (s)he is moved to act for the fun or challenge rather than for the external pressures or reward. That is, learners with intrinsic value pursue enjoyment of learning, understanding of new things, and therefore tend to regulate themselves in terms of cognition and behavior (Pintrich & DeGroot, 1990). Since learners in corporate context tend to enroll in e-learning programs on a needs-basis to improve their performance rather than because of external reward, intrinsic value is considered as a better predictor for the learning outcome (Ames & Archer, 1988). Therefore, intrinsic value was included as one of the factors in the present research model. In a previous study by Pintrich and DeGroot (1990), intrinsic value was highly correlated with the level of cognitive engagement of seventh graders in science and English classes. In addition, Wolters (2004), in a study conducted with 525 junior high school students, reported that students’ mastery orientation, which means intrinsic value, was a significant predictor of cognitive engagement, when added in a model with other predictors such as prior standardized achievement, gender, performance-approach structure, performance-approach orientation, performance-avoidance orientation, and self-efficacy.

While self-efficacy and intrinsic value are related to learner characteristics, usefulness and ease of use of the online learning programs are considered as important factors for learning. According to Davis (1989), as introduced as the part of the Technology Acceptance Model (TAM), perceived usefulness is defined as the degree to which a person believes that using a particular system will enhance his or her job performance. Perceived ease of use, on the other hand, refers to the degree to which a person believes that using a particular system will be free of effort. In an educational context, the particular system in Davis’ definition is substituted for learning program. A body of literature related to perceived usefulness and ease of use reported that the learner’s chance of experiencing flow during learning increases as the learners’ perceived usefulness and ease of use increase. Chang and Wang (2009) conducted a study to understand users’ communication behavior in Computer-Mediated Environments, employing the TAM. The result revealed that flow experience was affected by the perceived ease of use and the interactivity of online communication. Skaberg and Kimmel (2003) conducted a study applying the flow model to empirically evaluate visitors’ experience while browsing a web site. As a result, perceived ease of use indirectly affected flow experience, mediated by the interactivity of the web site.

In sum, a review of the literature indicated that learners’ self-efficacy, intrinsic value, and their perceived usefulness and ease of use may predict the experience of flow during online learning. Especially, self-efficacy and intrinsic value were suggested together as motivational components of learners’ self-regulated learning by Pintrich and DeGroot (1990). In the present study, these two learner variables, along with two external variables, usefulness and ease of use of e-learning program, are formulated as the research hypothesis.

**Factors related to achievement: Self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, and learning flow**

Among the many variables related to achievement, the following are discussed as meaningful predictors in this study: self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, and learning flow. Self-efficacy has been consistently reported as an influential factor. Judge, Jackson, Shaw, Scott and Rich (2007) conducted a meta-analysis to estimate the contribution of self-efficacy to work-related performance. The results revealed that self-efficacy predicted performance in jobs or tasks of low complexity. Especially, self-efficacy had an
indirect effect on job performance, when mediated by individual personality. Martin (2009), in a large-scale, correlational study on secondary and undergraduate students’ motivation in Australia, concluded that self-efficacy significantly correlated to learning achievement. Another study by Gore (2006) reported that undergraduate students’ academic self-efficacy was a significant predictor of learning outcomes such as GPA and enrollment status. In a corporate online training context, Joo, Kim and Kim (2008) conducted a study identifying factors affecting learning achievement. After collecting survey data from 1,130 adult learners, they concluded that self-efficacy, along with self-regulated learning skills and task value, predicted achievement significantly.

Previous studies also showed that achievement tends to be predicted by intrinsic value. Since Bloom (1983) claimed that students learn better when they are internally motivated, the role of internal value and/or goal-orientation has been discussed in ample research. Pintrich and DeGroot (1990) conducted a correlational study examining relationships between motivational orientation, self-regulated learning, and classroom academic performance, and concluded that self-efficacy and intrinsic value were positively related to cognitive engagement and performance. More recently, Spinath, Spinath, Harlaar, and Plomin (2006) conducted a study with 1,678 nine-year-old UK elementary school children taking part in the Twins Early Development Study. They reported that students’ intrinsic value contributes to the prediction of achievement in mathematics and English.

In addition to self-efficacy and intrinsic value, test anxiety is considered as another motivational factor influencing achievement (Pintrich & DeGroot, 1990). Test anxiety is likely to hinder concentration on performance, as it is defined as the experience of evaluating apprehension during the learning and exam process (Spielberger & Vagg, 1995). Mandler and Sarason (1952) examined the relationship between the level of text anxiety and science test scores of 186 middle school students, and reported that students with higher test anxiety scored lower than the students with lower test anxiety. For adult learners, Cassady and Johnson (2002) conducted a correlational study with 417 undergraduate students, and reported similar results with the level of test anxiety being negatively correlated with achievement scores. Although prior studies do exist in the traditional classroom environment, text anxiety in online learning has not been discussed sufficiently. Based on the framework suggested by Pintrich and DeGroot (1990), who incorporated test anxiety as one of the motivational factors along with self-efficacy and intrinsic value, this study employed test anxiety as a third predictor variable for achievement. Test anxiety is expected to explain students’ affective or emotional reactions to the task, which would provide more comprehensive understanding of online learning. The framework of Pintrich and DeGroot (1990) has been recognized as a meaningful model predicting academic performance (Eccles & Wigfield, 2002).

Perceived usefulness and ease of use of the online learning programs, as an external variable, is another influential factor for academic achievement. Johnson, Hornik and Salas (2008) conducted an empirical study to identify factors for successful e-learning in a university-level context. Based on the results of structural equation modeling, the study suggested that perceived usefulness was related to course performance and course satisfaction. Arbaugh and Duray (2002) also reported that perceived usefulness and perceived ease of use in web-based MBA programs were significant predictors of learning outcome as well as learners’ satisfaction. These results are not surprising, and have been supported by many researchers (e.g., Liaw, 2008; Roca, Chiu & Martinez, 2006).

Lastly, learning flow has been related with academic achievement in prior research. From a theoretical standpoint, Kiili (2005) developed a participatory multimedia learning model which is rooted in multimedia learning principles and learning flow, and claimed that learning activities requiring less cognitive resources tend to enhance the experience of learning flow, which will eventually produce better academic performance. Several studies have also demonstrated a positive correlation between learners’ engagement and achievement-related outcomes for elementary, middle, and high school students (Connell, Spencer, & Aber, 1994; Marks, 2000; Skinner, Wellborn, & Connell, 1990). Nystrand and Gamoran (1991) claimed that substantive engagement (similar to cognitive engagement) in the classroom was positively related to scores on an achievement test developed to measure students’ in-depth understanding and synthesis. As such, previous studies have implied that the experience of learning flow would help student focus on learning and demonstrate better achievement, even if the difficulty level of the tasks is quite high.

To summarize, this literature review has suggested that self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, and learning flow might play substantial roles in relation to academic achievement. Hence, these variables were included in the research model, which was formulated as a research hypothesis.
Mediating effect of learning flow

As the research model is conceptualized based on the prior research, learning flow was set as a mediator variable to connect the predictors - self-efficacy, intrinsic value, and perceived usefulness and ease of use - and achievement. Although not many empirical studies examining the mediating effect of learning flow have been conducted, the present study propose that self-efficacy, intrinsic value, and perceived usefulness and ease of use will impact achievement, mediated by learning flow. Theoretically, Baron and Kenny (1986) stated that if A influences B, and B influences C, then there may exist a mediating effect of B between A and C. Therefore, a hypothesis regarding the mediating effect of learning flow has been incorporated into this study.

Purpose of the study and research model

After an extensive literature review, the present researchers found that previous studies investigated either learners’ characteristics and motivation or learning environment issues, rather than incorporating these two into a comprehensive model. In addition, methodologically, correlational analysis and multiple regression analysis were most frequently adopted in the prior research (Harroff & Valentine, 2006; Morris, Wu, & Finnegan, 2005), which limits the interpretation of implications. This study intended to provide an integrated view in terms of research variables as well as methodological approach. Regarding research variables, self-efficacy, intrinsic value, and test anxiety were selected as learners’ motivational factors, while perceived usefulness and ease of use were also selected as learning environmental factors. Learning flow was considered as a mediator of predictors and achievement. Regarding methodological approach, structural equation modeling was employed in order to provide cause-and-effect inferences.

The purpose of this study was to examine the structural relationships among self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use of the e-learning courseware, learning flow and achievement. Representation of the hypothesized model tested in this study is shown in Figure 1.

This hypothetical model, derived from the literature review, was used to test the following hypotheses:

Hypothesis 1: Self-efficacy, intrinsic value, perceived usefulness and ease of use have positive effects on learning flow in the corporate e-learning environment.

Hypothesis 2: Self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, and learning flow have positive effects on achievement in the corporate e-learning environment.

Hypothesis 3: Learning flow has a mediating effect between predicting variables and achievement in the corporate e-learning environment.

Methodology
Participants

Participants in this study were the learners who were enrolled in e-learning programs in October 2009 at a large company in Korea. This company, running 30 sister-companies and 130 foreign branches, was selected because of its twelve-year history of implementing job-task-related e-learning courseware across the organization, which was expected to minimize the novelty effect of e-learning interventions to the learners. In addition, it was possible to examine the factors affecting the learning outcomes, since the learners shared an identical course registration system, learning management system, and evaluation criteria (Shea, Li, & Pickett, 2006). Two different surveys were administered for this study, with 326 and 271 learners responding to each. Of the 263 learners who responded to both surveys, 15 were eliminated due to incomplete responses. As a result, the study analysis was based on the remaining 248 usable responses. Demographically, there were more males than females (86.7% male, 13.3% female), and their age ranged from 23 to 58 (19.0% in their twenties, 52.8% in their thirties, 25.0% in their forties, and 3.2% in their fifties).

Measurement instruments

Several instruments, used or adapted from a variety of existing instruments, provided the study data. Table 1 presents the original sources, number of items implemented, and Cronbach’s alpha calculated after modification to suit the research context.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cronbach’s alpha</th>
<th># of items</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-efficacy</td>
<td>.90</td>
<td>9</td>
<td>Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich &amp; DeGroot, 1990)</td>
</tr>
<tr>
<td>Intrinsic value</td>
<td>.89</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Test anxiety</td>
<td>.87</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness and ease of use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Usefulness</td>
<td>.90</td>
<td>4</td>
<td>Technology Acceptance Model (TAM) (Davis, 1989)</td>
</tr>
<tr>
<td>- ease of use</td>
<td>.81</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

The instruments measuring self-efficacy, intrinsic value, and test anxiety were adopted from the Motivated Strategies for Learning Questionnaire (MSLQ) developed by Pintrich and DeGroot (1990). In order to assess self-efficacy, a 9-item, 5-point Likert scale was used, with 1 indicating ‘strongly disagree’ and 5 ‘strongly agree.’ Sample items are ‘I’m certain I can understand the ideas taught in this course’ and ‘Compared with others in this class, I think I’m a good student.’ Cronbach’s alpha from the present data was .90. Also, the construct reliability was .87, and the average variance extracted (AVE) was .92, demonstrating good convergent validity and discriminant validity. The instrument measuring intrinsic value consisted of 9 items using a 5-point Likert scale. Sample items are ‘It is important for me to learn what is being taught in this class’ and ‘Even when I do poorly on a test I try to learn from my mistakes.’ Cronbach’s alpha from the present data was .89. The construct reliability and AVE were .97 and .94, respectively. Test anxiety was measured using a 4-item, 5-point Likert scale. A sample item is ‘I am so nervous during a test that I cannot remember facts I have learned,’ and the Cronbach’s alpha from the present data was .87. The construct reliability and AVE were .89 and .80, respectively.

TAM suggested by Davis (1989) was employed to measure perceived usefulness and ease of use of the e-learning courseware. The instrument originally developed by Davis (1989) was translated into Korean by the present researchers and reviewed by two experts in educational technology field. There were 4 items for usefulness, and another 4 for ease of use. Sample items are ‘This e-learning courseware improved my job performance’ and ‘Learning to use the e-learning courseware was easy for me.’ Cronbach’s alpha for perceived usefulness and perceived ease of use were .90 and .81, respectively. The construct reliability and AVE were .92 and .86, respectively, demonstrating good convergent validity and discriminant validity.
The instrument used to measure learning flow for this study was the Flow State Scale (FSS), which was originally developed by Jackson and Marsh (1996), and subsequently validated by Martin and Jackson (2008). Nine items on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree), were included in the instrument. A sample item is ‘I am not concerned with what others think while I study.’ Cronbach’s alpha from the present data was .92. The construct reliability was .95, and the AVE was .91.

Achievement was measured by the scores from the final exam, consisting of 20 closed-ended items, which were randomly selected from the item pool. Learners were allowed to submit their answers only once. The possible achievement ranged from 0 to 60.

Data collection

Two online surveys were administered in order to analyze structural relationships among self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, learning flow, and achievement in corporate online learning. The first survey, collecting data related to self-efficacy, intrinsic value and test anxiety, was distributed during the first week of the program. The second survey, measuring perceived usefulness and ease of use, and learning flow, was delivered in the last week of the program. Achievement scores were collected from the database of the learning management system.

Data analysis

The hypothesized research model illustrated in Figure 1 was specified as the statistical model using latent variables (see Figure 2). Item parcels were used to minimize any possible overweight on a particular variable in the model, given that self-efficacy, intrinsic value, test anxiety, and learning flow are unidimensional factors. A parcel can be defined as an aggregate-level indicator comprised of the sum or average of two or more items (Kishton & Widamn, 1994), which is likely to reduce measurement error by using fewer observed variables and to ensure the assumption of multivariate normality (Bandalos, 2002; Sass & Smith, 2006). Multivariate normality was checked using AMOS 6.0 by observing the skewness and degree of kurtosis for all the measured variables together. Since each variable was normally distributed, maximum likelihood estimation was selected as an appropriate statistical estimation method. The goodness of fit indices used for this study were the minimum sample discrepancy (CMIN), Tucker-Lewis index (TLI), comparative fit index (CFI), and root-mean-square error of approximation (RMSEA), and the direct effects among the variables were tested at the significance level of .05.

Results

Descriptive statistics and correlations among the variables

The means, standard deviations, skewness and kurtosis for all the measured variables were analyzed together to check the normality assumption. The ranges of these statistics were 2.56 to 33.75, .50 to .10.15, .03 to .71 (absolute values), and .01 to 3.34, respectively (See Table 2). According to Kline (2005), absolute skewness values less than 3 and absolute kurtosis values less than 10 meet the assumption of the multivariate normal distribution of the data for structural equation modeling. Correlations were also examined to check the strength of the relationships among the variables of interest, and the results revealed significant correlations among all of the variables at the alpha level of .05 (see Table 2).

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Self-efficacy 1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Self-efficacy 2</td>
<td>.78*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Intrinsic value 1</td>
<td>.63*</td>
<td>.61*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Intrinsic value 2</td>
<td>.60*</td>
<td>.56*</td>
<td>.78*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Test anxiety 1</td>
<td>-.30*</td>
<td>-.30*</td>
<td>-.29*</td>
<td>-.26*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test anxiety 2</td>
<td>-.35*</td>
<td>-.29*</td>
<td>-.33*</td>
<td>-.33*</td>
<td>.75*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Means, standard deviations and correlation coefficients
### Assessment of measurement model

Based on the result of maximum likelihood estimation, Table 3 shows the goodness of fit indices for the *a priori* measurement model, indicating that the measurement model exhibited a good fit with the data collected (e.g., RMSEA=.000).

**Table 3. Fit statistics for the measurement model**

<table>
<thead>
<tr>
<th></th>
<th>CMIN (χ²)</th>
<th>p</th>
<th>df</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA (90% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement model</td>
<td>22.19</td>
<td>.625</td>
<td>25</td>
<td>1.003</td>
<td>1.000</td>
<td>0.000 (.000 – .044)</td>
</tr>
</tbody>
</table>

Note a: n=248

Note b: df = degree of freedom; TLI = Tucker-Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation.

![Diagram showing factor loadings](image)

**Figure 2. Result of confirmatory factor analysis**

According to Figure 2, the factor loadings ranged from .77 to .97, indicating the adequate validity of all the factors in the measurement model since all the loadings were greater than .50 (Hair, Anderson, Tatham, & Black, 1992). In
addition, the cross-loadings of the variables ranged from -.21 to .76, confirming that all constructs in the estimated model fulfilled the condition of convergent, as well as discriminant, validity. Therefore, the measurement model appeared to fit the data well and did not need to be changed.

**Structural model and hypothesis testing**

The proposed relationships were analyzed, as shown in Table 4, and the initial structural model provided a good fit to the data (e.g., TLI= 1.005; CFI= 1.000; RMSEA=.000(.000 ~ .038)).

<table>
<thead>
<tr>
<th>CMIN (χ²)</th>
<th>p</th>
<th>df</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA (90% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural model</td>
<td>26.51</td>
<td>.696</td>
<td>31</td>
<td>1.005</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: n=248

In order to test the hypothesis, direct effects among self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, learning flow, and achievement were examined at the alpha level of .05, by reviewing the β weights. First, the effects of self-efficacy, intrinsic value, and perceived usefulness and ease of use on learning flow were investigated. The effect of self-efficacy on learning flow was $\beta = .205$ (t= 2.203, p<.05), and the effect of intrinsic value on learning flow was $\beta = .204$ (t= 2.174, p<.05). The effect of perceived usefulness and ease of use on learning flow was examined as $\beta = .466$ (t= 7.565, p<.05). Second, the effects of self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, and learning flow on the achievement were investigated. The effect of self-efficacy on achievement was $\beta = -.122$ (t= -1.038, p>.05), while the effect of intrinsic value on achievement was $\beta = .232$ (t= 1.983, p<.05), the effect of test anxiety was $\beta = .152$ (t= -2.122, p<.05), and the effect of perceived usefulness and ease of use was $\beta = .171$ (t= 2.036, p<.05). Lastly, the effect of learning flow on achievement was $\beta = .078$ (t= .842, p>.05). To summarize, all of the direct effects were statistically significant other than the effects of self-efficacy and learning flow on achievement. Hence, the two insignificant path coefficients, “self-efficacy→achievement” and “learning flow→achievement”, were removed from the structural model to keep the model concise, as shown in Figure 3.

**Figure 3. Modified structural model**
As the original and modified models were in a hierarchical relationship, chi-square statistic was used to examine the statistical difference between the two models. The absence of any significant difference between the two models in terms of the goodness of fit ($\chi^2_D=5.06$, $p=.08$) confirmed the modified structural model as the final research model. Table 5 presents the goodness of fit indices for the modified model, indicating that the model exhibited a good fit (TLI=1.006; CFI=1.000; RMSEA=.000). Although the CMIN value of the modified structural model was 1.52 larger than that of the initially hypothesized model, the difference was not significant as mentioned above. The standardized path coefficients of the modified model are shown in Figure 4.

Table 5. Fit statistics for the structural model (modified)

<table>
<thead>
<tr>
<th>Model</th>
<th>CMIN ($\chi^2$)</th>
<th>p</th>
<th>df</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA (90% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural model (modified)</td>
<td>28.03</td>
<td>.713</td>
<td>33</td>
<td>1.006</td>
<td>1.000</td>
<td>.000 (~.000-.036)</td>
</tr>
<tr>
<td>Hypothesized model</td>
<td>26.51</td>
<td>.696</td>
<td>31</td>
<td>1.005</td>
<td>1.000</td>
<td>.000 (~.000-.038)</td>
</tr>
</tbody>
</table>

Note: n=248

Figure 4. Modified model with standardized path coefficients

The direct effects among the variables included in the modified model were tested (See Table 6). First, the effects of self-efficacy, intrinsic value, and perceived usefulness and ease of use on learning flow were investigated respectively. The effect of self-efficacy on learning flow was $\beta=.205(t=2.216, p<.05)$, and the effect of intrinsic value on learning flow was $\beta=.204(t=2.191, p<.05)$. The effect of perceived usefulness and ease of use on learning flow was examined as $\beta=-.468(t=-7.581, p<.05)$. Second, the effects of intrinsic value, test anxiety, and perceived usefulness and ease of use on the achievement were investigated. The effect of intrinsic value on achievement was $\beta=-.168(t=2.237, p<.05)$, the effect of test anxiety was $\beta=-.140(t=-1.980, p<.05)$, and the effect of perceived usefulness and ease of use was $\beta=.204(t=2.918, p<.05)$. The results indicated that self-efficacy, intrinsic value, and perceived usefulness and ease of use had significant effects on learning flow, thereby supporting the first hypothesis. The second hypothesis was partly supported by the result, since intrinsic value, test anxiety, and perceived usefulness
and ease of use had significant effects on achievement. Among these, perceived usefulness and ease of use had a relatively larger effect on learning flow and achievement as well. The third hypothesis was rejected, as the effect of learning flow on achievement was not statistically significant.

<table>
<thead>
<tr>
<th>Table 6: Effect decomposition for the modified model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 6. Effect decomposition for the modified model</strong></td>
</tr>
<tr>
<td>Direct effects</td>
</tr>
<tr>
<td>Learning flow ← self-efficacy</td>
</tr>
<tr>
<td>Learning flow ← intrinsic value</td>
</tr>
<tr>
<td>Learning flow ← perceived usefulness and ease of use</td>
</tr>
<tr>
<td>Achievement ← intrinsic value</td>
</tr>
<tr>
<td>Achievement ← test anxiety</td>
</tr>
<tr>
<td>Achievement ← perceived usefulness and ease of use</td>
</tr>
</tbody>
</table>

*p<.05

**Discussion**

The structural relationships among self-efficacy, intrinsic value, test anxiety, perceived usefulness and ease of use, learning flow and achievement were analyzed in this study. First, self-efficacy, intrinsic value, and perceived usefulness and ease of use had statistically significant direct effects on learning flow. This indicates that learners are likely to experience flow during learning when they have higher self-efficacy and intrinsic value, and when they perceive the e-learning courseware to be useful and easy to use. Regarding the effect of self-efficacy on learning flow, the result supports the previous study by Puzziferro (2008) who claimed that self-efficacy predicted learning flow and performance. In terms of the effect of intrinsic value on learning flow, the result is consistent with the claim made by Pintrich and DeGroot (1990) and by Miltiadou and Savenye (2003) that learners’ intrinsic value is one of the powerful predictors of learning. Also, the positive effect of perceived usefulness and ease of use on learning flow echoes the suggestions made by Johnson and colleagues (2008) for creating successful online learning environment, and by Harroff and Valentine (2006), who indicated that the message design of online learning contents influences the learners’ learning flow.

Second, intrinsic value, test anxiety, and perceived usefulness and ease of use had statistically significant direct effects on achievement. This result is consistent with that reported in a previous study by Miltiadou and Savenye (2003). Also, test anxiety appeared to have a negative relationship with achievement, as reported by Cassady and Johnson (2002), and Kleijn et al. (1994). The result related to perceived usefulness and ease of use supports the claim made by Shin (2006) who demonstrated that both of perceived usefulness and perceived ease of use influence learners’ achievement. Johnson and colleagues (2008) also suggested that perceived usefulness is a significant predictor of learners’ satisfaction and achievement.

However, contrary to the previous studies, self-efficacy and learning flow failed to predict achievement. For example, Puzziferro (2008) reported that self-efficacy was a significant predictor of achievement for freshmen in online university, and Pajares (1996) also claimed that self-efficacy, as much as individual learning capability, was a critical factor for learning. The present researchers noticed that the previous studies only examined motivational factors, and excluded external factors such as usefulness and ease of use of the learning program. In other words, although self-efficacy was a significant predictor in a model without external variables, the result may differ when factors other than learner characteristics are added to the model.

Third, the result indicated that learning flow was not a meaningful predictor of achievement, which is not consistent with the results from the prior research that reported learning flow as an influential factor for learning outcome (Kim, 2005; Nystrand & Gamoran, 1991; Seok, 2008). This finding may be partially attributed to the study context. For example, Seok (2008) recruited elementary school students, while Kim (2005) worked with undergraduate and graduate-level students. Given that the present study was conducted in a corporate setting where the learners had to
continue performing their own job-tasks and could not focus exclusively on the learning itself, the analysis of learning flow may need to follow a different approach in this context.

Conclusion

The purpose of this study was to investigate the direct effects of self-efficacy, intrinsic value, test anxiety, and perceived usefulness and ease of use on learning flow and achievement in corporate e-learning. In addition, the indirect effects of self-efficacy, intrinsic value, test anxiety, and perceived usefulness and ease of use on achievement, as mediated by learning flow, were also examined. The findings suggested that self-efficacy, intrinsic value, and perceived usefulness and ease of use affected learning flow, while intrinsic value, test anxiety, and perceived usefulness and ease of use were significant predictors of achievement. The results revealed perceived usefulness and ease of use to be the most influential factor for both learning flow and achievement. This suggests that instructors and instructional designers need to employ strategies that increase learners’ self-efficacy and intrinsic motivation in order to facilitate learning flow. To improve academic achievement by providing internal motivation to the learners, learning tasks should be designed so that they are relevant and valuable to learners. Most of all, the design of the learning environment should be centered around learners so that every feature and function of the online system is useful and easy to use.

Finally, four implications for further research to broaden the understanding and address the study limitations are presented. First, the role of self-efficacy in a comprehensive model incorporating both learner characteristics and external factors should be re-examined to confirm the present study results. Second, learning flow in a variety of research contexts should be investigated. As reported in this study, learning flow in a corporate setting may have a different function compared to that in an academic setting. Third, the measurement of learning flow needs to be elaborated in further research. Since the self-reporting instrument used in this study may not reflect the actual level of immersive learning experience, observation or interview should be performed to increase the robustness of the methodology. Lastly, the present study results can only be applied to the Korean corporate e-learning setting. In order to generalize the model, future study should included samples selected from other countries or various learning contexts.

Acknowledgement

This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (20100027744).

References


Providing Adaptivity in Moodle LMS Courses

Marijana Despotović-Zrakić, Aleksandar Marković, Zorica Bogdanović*, Dušan Barać and Srdjan Krčo
University of Belgrade, Faculty of Organizational Sciences, 11000 Belgrade, Jove Ilića 154, Serbia // maja@myelab.net // aca@fon.rs // zorica@myelab.net // dusan@myelab.net // krco@myelab.net
*Corresponding author

ABSTRACT
In this paper, we describe an approach to providing adaptivity in e-education courses. The primary goal of the paper is to enhance an existing e-education system, namely Moodle LMS, by developing a method for creating adaptive courses, and to compare its effectiveness with non-adaptive education approach. First, we defined the basic requirements of a course adaptation and the steps required to implement it. Based on the collected data, we identified the characteristic groups of students and then classified all participating students into these groups. The final adaptation of courses was performed in the distance education system of the University of Belgrade, in the Laboratory for e-business, Faculty of Organizational Sciences. Findings of the experimental study showed that the students’ effectiveness and achievements in learning were higher when they attended courses adapted using the described method, in comparison to the non-adaptive e-learning courses.

Keywords
Adaptive e-learning systems, e-Learning course adaptation, Group personalization, Moodle LMS

Introduction
Learning management systems (LMS) are powerful integrated systems that support a number of activities performed by teachers and students during the e-learning process. Teachers use an LMS to develop web-based course notes and quizzes, to communicate with students and to monitor and grade student progress. Students use it for learning, communication, and collaboration.

One of a widely used open-source systems for learning process management is Moodle (Modular Object Oriented Developmental Learning Environment). Moodle LMS is in use for the last five years in the teaching process at the University of Belgrade, in the Laboratory for e-business of the Faculty of Organizational Sciences. Each school year, more than 700 students use this system on all the study levels (undergraduate, master, and PhD studies). E-learning courses include the following areas: Internet technologies, E-business, Computer simulation, Mobile computing, and Internet marketing. In addition, we use the concept of blended learning throughout the whole teaching process.

Adaptive e-learning systems
In most cases, LMS users belong to heterogeneous groups with different individual, sometime even adverse, characteristics and needs. The adaptation of e-education systems to an individual or to a group based on their characteristics, expectations, knowledge, and preferences of the students is the next step in the evolution of the e-learning systems.

The lack of adaptive learning environments or an environment with adaptive features is partly due to the “one-size-fits-all” concept in use today which presents the same static content to all students. Several studies suggested that this type of e-learning system organization were not successful. Therefore, the emphasis is moving towards learner-oriented platforms and putting student’s expectations, motivation, habits, learning styles, needs, etc. in the focus of interest.

According to 0, an e-learning system is considered to be adaptive if it is capable of: monitoring the activities of its users; interpreting these on the basis of domain-specific models; inferring user requirements and preferences out of the interpreted activities, appropriately representing these in associated models; and, finally, acting upon the available knowledge on its users and the subject matter at hand, to dynamically facilitate the learning process. Since the system behavior adapts to a person, this kind of adaptation is also called personalization. Thus, adaptive e-
A learning system can be described as a personalized system, which in addition to the content discovery and assembly, is able to provide an adaptive course delivery, an adaptive interaction, and an adaptive collaboration support.

AWBES (Adaptive Web Based Educational System) were developed upon the AHES (Adaptive Hypermedia Educational System) theoretical concepts. The web based educational system incorporates adaptivity and intelligence by building a model of individual learner’s goals, preferences, and knowledge. The system is then adapted to the learner’s needs by using this model throughout the course delivery. For more details on theoretical concepts and the realization of these systems in practice please see 0 0 0 0 0 0.

Providing adaptivity in e-education system

Adaptive e-learning systems focus on adapting the courses to the individual characteristics of the students. These systems are found to perform better than LMS. Adaptive quizzes, intelligent solution analyzers, adaptive class monitoring systems, and adaptive collaborative systems perform their respective functions more efficiently.

We should bear in mind, however, that these systems have a certain number of constraints:

- Creating adaptive e-learning systems and their incorporation into the educational processes is a rather complex and expensive process that requires a high level of engagement from all participants in the process.
- Such systems are designed for the purpose of ensuring adaptive functionality, however, in practice they frequently turn out not to fit into the learning process.
- One common problem is that the learning content is impossible to recreate and use anew. The systems are not interoperable. They are developed as isolated applications and they cannot exchange either resources or data on students. The research has shown that these systems are most often isolated applications and experimental solutions with a limited interface.
- The basic services of the e-learning systems, such as course administration, contents creating, etc. are much more difficult to carry out in adaptive systems.
- Due to the complexity of the systems, the users are expected to have some pre-existing knowledge of the system itself.
- The services that often lack most in the communication and social interaction between the participants in the e-learning process are chat, forum, and videoconference.
- For many courses, the time is an important factor, i.e. it is not possible to learn about the students’ needs during a course and then adapt it due to the lack of time. In such scenarios, it is preferred to have information about the students’ preexisting knowledge and experience as well as the intended usage of the course content in advance to maximize the time a student spend in a personalized environment. AWBES does enable full adaptation, but it does not guarantee its timeliness.

The key problems in adaptive systems, therefore, are reflected in the architecture and realization of the system, rather than in its essence – adaptivity. On the other hand, the LMS do not provide an adequate level of adapting of contents and services.

Some authors present new approach to adapting LMS: embed AH functionality into modern web-based courses using adaptive hypermedia services that can be chosen by the classroom teacher and attached to courses as a single reusable component. Adapting LMS in such a way requires development of new software modules. These software modules can become obsolete and incompatible with each new version of LMS which makes the process inefficient.

Personalization based on learning styles

Student modeling is the process whereby an adaptive learning system creates and updates a student model by collecting data from several sources implicitly (observing user’s behavior) or explicitly (requesting directly from the user). Traditionally, the majority of the student modeling systems have been limited to maintain assumptions related to student’s knowledge (acquired during evaluation activities) without paying too much attention to student’s preferences.
Learning is a cognitive activity that differs from a student to a student. Analyzing adaptability in an e-learning system has explicitly pointed out the importance of the modeling learners’ cognitive characteristics, particularly learning styles as the most explored cognitive features.

There are several different learning style models presented in literature; however, Felder-Silverman Learning Styles Model (FSLSM) is often used for providing adaptability regarding learning styles in e-learning environments. Felder-Silverman model describes a single student along four dimensions: 1) Active and reflexive learning style; 2) Sensitive and intuitive learning style; 3) Visual and verbal learning style; 4) Sequential and global learning style.

**Research objectives**

The primary goal of the study was to develop an efficient method for creating adaptive distance education courses in the Moodle LMS. The courses will be organized and adapted to groups of students. Adaptation is based on students’ characteristics, particularly learning styles. The following steps were used:

- Defining main phases and requirements in developing personalized e-education systems;
- Clustering students by their learning styles;
- Personalization and adaptation of e-learning courses.

The key idea of this paper is based upon adaptation of courses within the existing Moodle LMS, in order to eliminate the need for developing new modules or designing entirely new solutions. The e-learning system used in the research and described in this paper is used by a large number of users/students daily. The idea was to perform adaptation in such a way to make process economically justified, not time consuming, bearing in mind the environment in which the courses were conducted.

Personalization we would achieve in our system is a personalization for a group, not for a person. We claim that this approach is appropriate for three reasons:

1. When the courses are attended by a large number of students, due to the numerous differences in their individual characteristics, the personalization for each student individually is not justified. Taking into consideration the existing differences, some common features can be identified and then they can be assigned into respective groups on the basis of these features.
2. Group personalization can be achieved at the beginning of the course. Therefore, a student will be provided with an adapted online course during the whole learning process.
3. Group personalization is cheap and efficient. Teachers need to adapt learning materials and activities to typical groups of students instead of to each student. This is particularly efficient when a large number of students attend the course.

If the adaptation is done at the beginning of the learning process, some of the AHS disadvantages will be overcome. By using students’ learning styles it is possible to define typical learning groups and to assign new students into the groups at the beginning of their courses. The data needed for classifying the new students into these groups can be gathered promptly, by using an appropriate questionnaire or a test, and analyzed with some of the business intelligence tools. Each group of students will be provided with specific teaching materials and activities, suited to their learning style. This approach enables adaptation controlled by teacher, and also adaptation that does not depend upon the version of Moodle LMS. It is not necessary to change or add modules. The adaptation is achieved through the customization of Moodle LMS. Personalization is not performed for all the characteristics, but only for those the research findings designate as the most important. It is in such a way that the economic and time efficiency of the adaptation process is achieved.

In the proposed model, there are no automated adaptation mechanisms. However, the activities the course organizers have to perform to carry out the course adaptation can be automated to a certain extent.

**Method for Moodle course adaptation**

In order to improve the process of personalization in e-learning systems and improve its effectiveness, its main phases and requirements have to be identified first. The steps shown in Figure 1. should not be conducted
separately, but as integrated components of an iterative and dynamic process of developing an adaptive e-education system.

Data for this analysis come from a testing sample of 318 undergraduate students at the University of Belgrade, Faculty of Organizational Sciences. Experimental group consisted of 218 and control group consisted of 100 students.

**Methods and procedure**

1. **Defining e-personalization goals and models:**

   The main goal of this study was to implement an adaptive e-learning system. The substantial part of the research was realized through dividing students who attend our e-learning courses into different clusters according to their learning styles.

2. **Collecting data about students for course personalization**

   Data about the students are collected from two sources: the questionnaire and the introductory course.

   **Phase - Questionnaire**

   Due to the personalization process constraints, not only in relation to the courses, but also to the distance education system itself, we designed a questionnaire according to the FSLSM. Main constraints are related to: time limit (only one semester for course), not having enough data about all students’ characteristics, big number of students, variety of different topics, materials, and e-learning course activities. In order to overcome these problems it was decided to formulate the questions so as to represent the four FSLSM dimensions of learning styles. The FSLSM dimensions are then connected to the functionality of the Moodle to enable easier implementation in a real system.
The survey consisted of 30 questions of which some dealt with general information (average grade, year of study) while the others were about motivation for learning, preferred style of communication, the manner of presenting subject matter, managing time and team work.

II phase – Introductory course

In order to obtain more data about students’ learning styles, we organized an introductory course that lasted for one week, involving all types of activities, resources, and materials. The course was created in Moodle LMS. It dealt with different aspects of e-business: e-commerce, e-government, customer relationship management, Internet marketing. The course was not adapted in any way, just like the other courses in Moodle. In order to complete the course, students had to pass a test designed to assess knowledge acquired from different types of materials and activities: text and multimedia materials, web pages, as well as activities available in Moodle (assignments, lessons, quizzes, forum discussions, etc). The test results were to be used to determine students’ learning styles.

3. Exploring collected data

All the data collected in the two described phases were integrated into a single table in such a way that each question from the questionnaire and from the test represented one column. The rows represented answers of the students. This table was suitable for further analysis and data mining. Upon integration, the data were transformed and reduced. The number of options in the answers were transformed and reduced to three, and the missing data were replaced with mean values.

4. Classifying students

Data mining is one of the business intelligence techniques and can be defined as the nontrivial extraction of implicit, previously unknown and potentially useful information from large data sets or databases. Although personalized recommendation approaches that use data mining techniques have been first proposed and applied in e-commerce for product purchase, different data mining techniques were also applied within the e-learning recommender systems.

By using data mining tools and techniques it is possible to conduct an intelligent analysis of large quantities of data stored in a database. Data mining can be used both as a means for predicting unknown or future values of the attributes of interest and for describing embedded patterns that will contribute to generating the best possible personalized e-learning models.

Clustering, as a data mining technique, was applied in building data mining model on the integrated data set prepared in the previous phase. Clustering algorithm finds natural groupings among data related to sets of input attributes, so that attributes within one group (cluster) have approximately the same values, while notable differences exist between groups (clusters). It could be asserted that the main aim of clustering is to discover hidden values and variables, upon which data can be precisely arranged. Clustering was carried out using the SQL Server Analyses services, a Microsoft clustering algorithm based on the K-mean algorithm.

A set of guiding questions were formulated in order to lead the process of creating data mining structures in the right direction:

- What is the most appropriate number of clusters?
- What are the main characteristics within the clusters and what are the differences between them?
- Which input variable, i.e. learning style, has a dominant influence in grouping the students?
- Is the mining model created suitable for making forecasts?
- Which content should be delivered to students from individual clusters?

The answers to these questions are given in the text that follows.

An experiment was conducted for the cases of two and three clusters. By processing and mining available data, it was determined that the results showed almost the same level of accuracy, regardless of whether the students were divided into two or three clusters. However, in the latter case, the outcomes were more consistent, logical and of higher quality. Therefore, the outputs and the conclusions drawn from the experiment with three groups will be presented here.
For each cluster the most important characteristics and related learning styles are presented in Figure 2.

![Figure 2. Clusters and related learning styles](image)

Distribution of students by clusters is shown in Table 1. Most of the students were classified into cluster 2, which represents a combination of intuitive, active and global learning styles.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>% of students in the cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>26%</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>61%</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>13%</td>
</tr>
</tbody>
</table>

Figure 3 shows factors that have the biggest impact on classification. The aptitude for the presentation of teaching materials is of greatest importance for the classification of students. Also, it is important who sets the topics for essays and deadlines for exam obligations, the teachers or the students themselves.

![Figure 3. Factors that have the biggest impact on classification](image)
Created data mining model needs to be validated. The comparison method applied here is called ‘the mining accuracy lift chart’ (Figure 4). The results were sorted and plotted on the graph together with the ideal model, that is, a theoretical model with the accuracy of 100%.

If the state of the predictable column, i.e. forecasted class in this case, is specified, than the quality of the model can be analyzed. The X-axis of the chart represents the percentage of the test dataset that is used to compare the predictions. The Y-axis of the chart represents the percentage of values that are predicted to be in the specified state, i.e. Cluster 1. It is important that it is significantly above the blue line that represents “random guess line”. The line that indicates the “ideal model” shows that it would “catch” 100% of predicted population (Cluster 1) by using about 50% percentage of the available data in the model with three clusters. The classifying model would catch 100% of students arranged in Cluster 1 by using approximately 80% of total data.

![Figure 4. Model validation - lift chart](image)

It is essential to point out that the clustering results obtained served as the basis for the adaptation.

5. **Courses adaptation according to the defined groups (clusters)**

Using the results of this research, the students were divided into three groups and changes and adaptations have been made in our e-learning courses for these 3 groups. The courses were undertaken in accordance with the activities, the scope and the schedule proposed in the previous phase. The next paragraph describes in detail how the adaptation was done.

There were some similar characteristics for all students, like working in teams or passing exams sequentially. Indicated requirements were filled for all of them by making some options globally available. To build the initial model, the system’s authors first must establish the rules to match learning styles with the resource’s characteristics in order to determine which resources are better suited to a particular learning style. According to Moodle, Moodle is LMS that can provide all important functionalities and services. At the same time, Moodle is flexible when it comes to insertion of new elements and integrations with other systems and technologies necessary for the adaptation to take place. Table 2. shows the relations among different learning styles and activities in Moodle learning management system. The adaptation of courses is performed in accordance with the presented rules.

6. **Adaptive courses realization**

Due to the fact that the main goal was to do some fine-tuning of the courses, the final adaptations based on proposed generic model and discussed relationships between learning styles and different types of presentation were reflected in:
• Course level adaptation - students in Group 1 and 2 were allowed to determine time-limits for finishing the examination, while certain conditions were set for the students in Group 3.
• Teaching materials - Group 1 was provided with pictures, videos, graphs, animation, and hypertext materials; Group 3 was provided with text and audio materials. Group 2 was provided with combination of multimedia and written materials.
• Examination - Students in Group 1 were assigned to do projects or case studies. At the same time, students in Group 3 were obliged to do essays instead of projects. Learners from Group 2 could choose between practical and theoretical tasks.
• Activities - Generally, almost all activities were available for students. The only difference was the way these activities were organized in (Figure 5.).

<table>
<thead>
<tr>
<th>Moodle activities</th>
<th>Collaboration methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forum</td>
<td>Chat</td>
</tr>
<tr>
<td>Active</td>
<td>Concrete problems</td>
</tr>
<tr>
<td>Reflexive</td>
<td>Topics for thinking</td>
</tr>
<tr>
<td>Visual</td>
<td>No</td>
</tr>
<tr>
<td>Verbal</td>
<td>Yes</td>
</tr>
<tr>
<td>Sequential</td>
<td>Yes</td>
</tr>
<tr>
<td>Global</td>
<td>Global topics</td>
</tr>
<tr>
<td>Sensitive</td>
<td>Facts, examples</td>
</tr>
<tr>
<td>Intuitive</td>
<td>Abstract topics</td>
</tr>
</tbody>
</table>

7. Assessment of personalized courses

The final phase in developing an adaptive e-learning course is the testing of its effectiveness, functioning of the personalized model, and comparing it with the effectiveness of the course before adaptation. Experiment included three instruments: a) pre-test, b) post-test, and c) survey for examining students attitudes towards adapted online courses. Students in both experimental and control group have taken pre-test after the introductory course. Pre-test
Results and discussion

In this chapter we present and discuss the findings of the experiment. Table 3. shows the percentage of students who passed or failed the exam in e-business, organized through nonadaptive and adaptive online courses. The number of students that passed exam by attending a nonadapative course is close to the average number from the previous years. The percentage of students who passed the exam is by 11% higher in the case of the adaptive e-learning environment.

Table 3. Students’ performance in the non-adaptive and in the adaptive course

<table>
<thead>
<tr>
<th></th>
<th>Adaptive course</th>
<th>Nonadaptive course</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of students who passed</td>
<td>94%</td>
<td>83%</td>
</tr>
<tr>
<td>% of students who failed</td>
<td>6%</td>
<td>17%</td>
</tr>
</tbody>
</table>

Comparing the students’ marks on the non-adaptive course with those on the adaptive course, we can see that it is in the adaptive course that a larger number of students got higher marks. In the case of the non-adaptive course, students’ marks were distributed near to normal distribution, while in the adaptive course a fairly large number of students received the highest mark (Figure 6.). In the grade system used, mark 5 is assigned to students who failed, while the marks of the students who passed range from 6 (lowest passing mark) to 10 (highest mark).

Statistical analysis of the gathered data was performed in order to determine if results achieved by students are statistically significant. We used:

- Independent t-test to determine if the difference between experimental and control groups was statistically significant in the results achieved in the pre-test.
- Independent t-test to determine if the difference between experimental and control groups was statistically significant in the results achieved in the post-test.
- Paired samples t-test to determine if the difference in the results in the pre-test and the post-test for students in the experimental group is statistically significant.
- Paired samples t-test to determine if the difference in the results in the pre-test and the post-test for students in the control group is statistically significant.
We defined the following statistical hypotheses:
- Null hypothesis H1: There is no significant statistical difference in the pre-test results between the experimental and the control group
- Null hypothesis H2: There is no significant statistical difference in the post-test results between the experimental and the control group
- Null hypothesis H3: There is no significant statistical difference between the pre-test and the post-test results of the experimental group
- Null hypothesis H4: There is no significant statistical difference between the pre-test and the post-test results of the control group

The comparative descriptive statistics is shown in Table 4.

Table 4. Comparative descriptive statistics of pre-test and post-test results

<table>
<thead>
<tr>
<th>Results</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
</tr>
<tr>
<td>Experimental group</td>
<td>218</td>
<td>218</td>
<td>8.05</td>
</tr>
<tr>
<td>Control group</td>
<td>100</td>
<td>100</td>
<td>8.14</td>
</tr>
</tbody>
</table>

The results of the independent t-test showed that the difference between the experimental and control group results achieved in pre-test is not statistically significant ($t=-0.706$, $df=316$, $p>0.05$), which supports the null hypothesis H1. The results of the independent t-test also showed that difference between the experimental and control group results achieved in post-test is statistically significant ($t=1.700$, $df=316$, $p<0.05$), therefore, we reject the null hypothesis H2.

Using the results of the independent t-test, we can conclude that students in the experimental and the control groups achieved the same results in the pre-test, but the results achieved in the post-test are better in the experimental group.

To compare the results of the pre-test and the post-test, we have used a paired samples t-test. The paired samples t-test showed that the difference in the results achieved by the students in experimental group in the pre-test and the post-test is statistically significant ($t=-0.202$, $df=217$, $p<0.05$), therefore, we reject the null hypothesis H3. The paired samples t-test also showed that difference in the results achieved by the students in the control group in the pre-test and the post-test is not statistically significant ($t=0.708$, $df=99$, $p>0.05$), which supports the null hypothesis H4.

The results of the paired samples t-test show that the students in the experimental group achieved significantly better results in the post-test in comparison with the pre-test, whereas the students in the control group did not achieve significantly better results in the post-test.

Finally, we examined the students’ attitudes towards the adapted e-learning system. The students were asked to answer each of the questions in the survey in that they could choose one of three options: “strongly agree”, “agree” and “disagree”. The survey consisted of three questions:
- Teaching materials and type of presentation are adapted to my learning style.
- Activities in the online course are adapted to my learning style.
- I have a positive attitude towards adapted online course.

Since the type of presentation proved to be of primary importance for the classification of students into clusters, we asked the students if the presentation of the subject matter in adapted courses was suited to their learning style. Only 5% of the students replied that the type of presentation was not adjusted to their learning style, while 50% agreed and 45% strongly agreed that the type of presentation was adapted to their learning style.

When asked if the adaptive course suited their learning style most students strongly agreed that it did (49%), 47% of students agreed, while only 4% answered negatively.

Finally, when asked about the overall impression about the adapted online course, 50% of the students reported that their attitude was positive (strongly agree), 45% of the students had a moderately positive attitude (agree), while only 5% of the students replied that their attitude towards the adapted course is not positive.
Conclusion

Conducted research showed that adaptation of e-courses leads to better results. Proposed method for course adaptation is based on adapting presentation of content, communication methods, and organization of online activities to students’ learning styles and preferences. Analyses of e-learning system adaptivity have shown that cognitive characteristics of students, such as learning style, are of the greatest importance for successful adaptation. Described method for course adaptation is based on using data mining techniques to classify students into clusters with regards to Felder-Silverman learning styles model. Type of presentation of teaching materials had the highest impact on clustering and classification. E-learning courses were adapted using relations between Felder-Silverman’s learning style model and activities in Moodle learning management system. Research results proved that teaching resources and activities adapted to learning styles lead to significant improvement in learning results.

Analysis of results showed that students who attended adapted online courses achieved better results than students who attended nonadapted online course. T-test showed that there was statistically significant difference in results achieved by students in experimental group on post-test compared to results of students in control group. Higher number of students passed the test, and they achieved higher grades.

Analyses of satisfaction of students in experimental group showed that they had positive attitude towards adapted online course, and they thought that course materials and activities had been adapted with respect to their learning styles.

One of the greatest advantages of described method for course adaptation is that it does not require programming new software. Adaptation can be performed without any programming knowledge, and teachers can create adapted courses by adjusting teaching materials and activities in Moodle LMS. The authors of this paper believe that this approach can easily be modified for application in other LMSs as well.

The biggest disadvantage of the described approach is that it does not enable real time adaptation. Therefore, it is necessary for teachers to monitor students’ progress and move them to another cluster, if necessary. This type of adaptation requires more teachers’ effort in both creating adaptive course and exploitation.

In this research, course personalization has been realized with respect to learning styles as only criterion. Future work will be directed toward providing further data about students with respect to different characteristics, such as pre-knowledge, expectations, etc. Future work will also be directed towards qualitative analysis of described method, as well as suitability for teaching other subjects and application in other learning management systems.

Acknowledgement

The authors of this paper are grateful to MNTRS for financial support grant no. 174031.

References


Agent Prompts: Scaffolding for Productive Reflection in an Intelligent Learning Environment

Longkai Wu and Chee-Kit Looi
National Institute of Education, Nanyang Technological University, Singapore // longkai.wu@gmail.com // cheekit.looi@nie.edu.sg

ABSTRACT
Recent research has emphasized the importance of reflection for students in intelligent learning environments. This study tries to investigate whether agent prompts, acting as scaffolding, can promote students’ reflection when they act as tutor through teaching the agent tutee in a learning-by-teaching environment. Two types of agent prompts are contrasted in this research, both from the perspective of a tutee, and they differ in their specificity. Generic prompts are aiming at eliciting students’ double-loop reflection on metacognitive strategies and beliefs. Specific prompts, on the other hand, are to encourage students’ single-loop reflection on domain-related and task-specific skills and articulation of their explanatory responses. Our findings suggest that designers of intelligent learning environment should concentrate on fostering students to reflect on their metacognitive strategies and beliefs, and allow students to take responsibility for directing their own learning autonomy.

Keywords
Question Prompts, Scaffolding, Learning-by-Teaching, Double-Loop Learning Theory

Introduction

Reflection, which is perceived as an active process of learning through experience (Dewey, 1910/1933), has been prominent in educational literature (e.g., Boud, Keogh, & Walker, 1985; Kolb, 1984; Moon, 1999). Kenmis (1985) suggests that reflection is a “dialectical process” that looks both inward at the individual’s thoughts and outwards at the situation and is a “meta-thinking” process where the relationship between thought and action in a particular context must be considered before moving to further thought and action. The ability to reflect has been associated with the higher levels of learning in a number of taxonomies of learning objectives (Biggs & Collis, 1982; Bloom, 1956). Moon (1999) considers the ability to carry out meaningful reflection as indicative of the highest level of deep learning. Chi, Siler, Jeong, Yamamichi and Hausmann (2001) have evidenced that only the number of student turns coded as “reflection”, which were comprehension-monitoring statements, is positively correlated with deep learning gains in the face-to-face (FTF) tutorial dialogues.

Question prompts, as suggested by previous studies (Davis & Linn, 2000; Ge & Land, 2004; Lai, 2008), can be an effective way to elicit reflection since they provide cognitively complex ways learners think about, feel about, and make connections in experience. Moon (2004) has suggested structuring reflection with questions to deepen the quality of reflection. Amulya (2004) notes that, by engaging in reflective activities such as responding to the question prompts, learners could build their understanding and locate the significance of their activities in a larger context. Thus they are enabled to observe the meaning they have taken from the experience and excavate the underlying qualities that made the experience significant. Specially, recent research shows the evidence of positive influence of tutee’s question prompts to tutor’s reflective learning in the context of peer tutoring (Cohen, Kulik, & Kulik, 1982; Graesser, Person, & Magliano, 1995; Roscoe & Chi, 2008). Roscoe and Chi (2008) find that tutee questions can motivate tutor explanations and metacognition, and thus have a significant and positive influence on the tutor’s learning activities and opportunities.

Researchers in ILEs (Intelligent Learning Environments) have also recognized the importance of incorporating question prompts into ILE designs (e.g. Hmelo and Day, 1999). Question prompts have been used as scaffolds to help direct students towards learning-appropriate goals, such as focusing student attention and modeling the kinds of questions students should be learning to ask (Azevedo and Hadwin, 2005). Positive evidences are found for question prompts to help students with various aspects, such as knowledge integration (Davis and Linn, 2000) and ill-structured problem-solving (Ge and Land, 2004; Xie and Bradshaw, 2008). Meanwhile, mechanisms for supporting self-explanation (Aleven & Koedinger, 2002), tutorial dialog (Graesser, VanLehn et al., 2001) or reflective dialog (Katz, O’Donnell & Kay, 2000) have been prevalent in traditional intelligent tutoring systems (ITS), in which the computer plays the role of tutor (e.g. Cognitive Tutor, AutoTutor). However, little research has addressed question prompts to help students towards learning-appropriate goals.
prompts as scaffolding strategy to elicit the learner’s reflection with a computer simulated tutee when playing the role of tutor within a learning-by-teaching context.

This paper is concerned with the investigation of agent prompts, i.e., question prompts to initiate learners’ reflection in learning from an agent-enabled learning-by-teaching environment (an adapted version of Betty’s Brain proposed by Biswas, Schwartz, Bransford, & TAG-V, 2001). We intend to explore how a learning partner, acting as the role of inquisitive tutee enabled by the generation of question prompts, might be used to address the challenge of facilitating reflection in a student tutor linked to learning-by-teaching experiences. The reflection of the student tutor mainly refers to an intermingled process of knowledge construction and metacognition as a direct result of his engagement in instructional activities inherent to the virtual tutoring process, such as explaining, answering questions from the tutee, correcting errors of the tutee and asking questions to the tutee (Cohen, 1986; Gartner, et al., 1971; King, 1998; Roscoe and Chi, 2007).

The remainder of the paper is organized as follows. First, we present the system design of agent prompts generator within the adapted Betty’s Brain. Second, the classroom study is described. Third, the results of how these agent prompts deliver their proposed benefits to secondary school students are provided. Lastly, discussions and conclusion for this work are described.

Agent Prompts Generator

Overview

The development work was focused on the generation and incorporation of meaningful agent prompts, which can arouse student’s reflective learning-by-teaching activities. The work was built on an existing system, Betty’s Brain (Figure 1), a learning-by-teaching agent environment built by the Teachable Agent Group in Vanderbilt University (Biswas, Schwarz, Bransford et al., 2001). With the ability to learn what the students have taught by concept mapping, Betty’s Brain was used to play the role of agent tutee and provide a context for students to practice tutoring in our research.

Aspects of Consideration

Based a reflection assistant model proposed by Gama (2004), three aspects were considered when the agent prompts generator was designed to produce appropriate question prompts to guide students in learning-by-teaching activities:
Learning-by-Teaching Stages, Reflection Types and Patterns in Student Maps (Figure 2). We discuss the three aspects respectively in the following sections.

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning-by-Teaching Stages</td>
<td>Familiarization</td>
</tr>
<tr>
<td></td>
<td>Production</td>
</tr>
<tr>
<td></td>
<td>Evaluation</td>
</tr>
<tr>
<td></td>
<td>Post-Task Reflection</td>
</tr>
<tr>
<td>Reflection Types</td>
<td>Double-Loop Reflection</td>
</tr>
<tr>
<td></td>
<td>Single-Loop Reflection</td>
</tr>
<tr>
<td>Patterns in Student Maps</td>
<td>Missing or Incorrect Expert Concepts</td>
</tr>
<tr>
<td></td>
<td>Missing or Incorrect Expert Propositions</td>
</tr>
</tbody>
</table>

*Figure 2. Aspects of Consideration in Designing Agent Prompts Generator*

**Learning-by-Teaching Stages**

The inquisitive agent tutee system offered a structured approach to help students proceed through the learning-by-teaching activities within four stages (Figure 3), which follow the conceptual stages in practicing tutoring, and provide different agent prompts through their actions in accordance with the structured stages.

![Learning-by-Teaching Stages Diagram]

*Figure 3. Learning-by-Teaching Stages*

As shown in Figure 3, the *Familiarization* stage is to let students to self-assess their understanding of domain knowledge and learning difficulties, as well as selecting their metacognitive strategies. The *Production* stage is to enable students to teach the agent tutee what they have learned by constructing concept maps and monitoring the agent tutee’s understandings. The *Evaluation* stage is to provide students with opportunities to evaluate the performance of the agent tutee, as well as their own performance. The *Post-Task Reflection* stage is to promote post-practice reflection on the learning-by-teaching experiences and the strategies being implemented.
Reflection Types

This study attempts to adopt and adapt the double learning theory proposed by Argyris and Schön (1974, 1978, 1996), which pertains to learning to change underlying values and assumptions, as the theoretical framework to support the system and experimental design. Such a theory includes two reflection types, single-loop reflection and double-loop reflection.

The single-loop reflection refers to increase efficiency of reaching an objective. It is task oriented and about the design of the process to retain reliability. It is simple reflection that may challenge assumptions and strategies to alter the plan of action but always ‘in ways that leave the values of a theory of action unchanged’ (Brockbank and McGill, 1998). Knight (1996) argues that on occasions there is use for this type of reflection because of a need for “developing and improving the realization of relatively fixed goals and objectives” (Knight, 1996, p. 13). Courtney et al. (1998) describe this as a lower level of reflection as it involves keeping to a set of rules and is simply error correction. This type of reflection is viewed valuable for day-to-day activities and is necessary for progress to be made within established frameworks (Brockbank and McGill, 1998).

The double-loop reflection is described by Courtney et al. (1998) as a higher level of reflection than single-loop reflection. This second loop could focuses on the examination and reflection of the theory or perspective in use (Lynch and Joham, 2004) or the evaluation of an experience using explicit and varied concepts (Lynch and Metcalfe, 2006). It is recognized that the action and consequences striven for in the first loop may not be valid – that there may be different perspectives regarding what the outcome should be and therefore assumptions, premises and context are questioned. Sterman (1994) has provided an excellent case for the contention that individual learning is enhanced through information feedback from a double-loop learning process in virtual worlds.

Considering from the perspective of double loop learning theory, we design agent prompts to elicit two major types of reflection for students in the learning-by-teaching environment.

- **Generic Prompts eliciting Double-Loop Reflection** are to lead students to examine their perspectives, assumptions and experiences by reflecting on metacognitive strategies and beliefs in learning and teaching. Sample generic prompts are: “What do you think about teaching and who is it for?”, “Before starting to teach, can you think about what you are supposed to learn from it?”, “Can you reread the learning objectives and resources and ask if the map really meets the description in the learning objectives and resources?”

- **Specific Prompts eliciting Single-Loop Reflection** are to lead students to reach certain learning objectives by reflecting on task-specific and domain-related skills regarding their activities and to articulate their explanatory responses. Sample specific prompts are: “Can you explain the concepts you just taught me?”, “Can you tell me if my reasoning process is correct and give me a further explanation?”

Patterns in Student Maps

Table 1 shows the sample patterns found in student maps and corresponding agent prompts stored in the repository. The algorithms of detecting patterns and fuzzy integrating expert maps used in our study are described in (Wu & Looi, 2008).

<table>
<thead>
<tr>
<th>Sample Patterns</th>
<th>Corresponding Agent Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Missing Expert Concepts</strong></td>
<td>Generic Prompts: “Can you review all things to check for missing important parts to teach me and give me an explanation?”</td>
</tr>
<tr>
<td>(e.g. Missing “Income” concept)</td>
<td>Specific Prompts: “Do you consider that you could teach me the concept of ‘Income’ and give me an explanation?”</td>
</tr>
<tr>
<td><strong>Incorrect Expert Concepts</strong></td>
<td>Generic Prompts: “Can you stop and review each part in the map to see if you have made a mistake and give me an explanation?”</td>
</tr>
<tr>
<td>(e.g. Incorrect “Electricity” concept)</td>
<td>Specific Prompts: “Do you want to reconsider the concept of ‘Electricity’ you teach me and give me an explanation?”</td>
</tr>
<tr>
<td><strong>Incorrect Expert Propositions</strong></td>
<td>Generic Prompts: “Can you stop and review each part in the map to see if a mistake has been made and give me an explanation?”</td>
</tr>
</tbody>
</table>
Methodology

Research Questions and Hypotheses

Our research questions focused on whether the systematic use of the agent prompts would be relevant to (a) the acquisition of learning outcomes and (b) the elicited levels of reflection.

The two research questions are as follows:
1. Do agent prompts have cognitive and metacognitive effects on secondary school participants’ learning outcomes in the domain of elementary economics within the learning-by-teaching environment?
2. Do agent prompts presented in different types (generic prompts vs. specific prompts) elicit different levels of reflection in their learning-by-teaching processes?

Based on the two research questions, we formulated the following hypothesis based on the assumption that double-loop learning is a higher level of learning than single-loop learning (e.g. Courtney et al., 1998).

**Hypothesis 1** (Learning Outcomes): The GP (Generic Prompts) group will achieve better learning outcomes than the SP (Specific Prompts) and the NP (No Prompts) groups, and the SP group will achieve better learning outcomes than the NP group. This hypothesis is it will be measured in three areas.

**Hypothesis 1.1** (Pretest/Posttest Gain): The GP group will show greater improvements from pretest to posttest than the SP and NP groups, and the SP group will show greater improvements from pretest to posttest than the NP group.

**Hypothesis 1.2** (Quality of Student Maps): In the main study, the GP group will develop better student maps than the SP and NP groups, and the SP group will develop better student maps than the NP group.

**Hypothesis 1.3** (Ability to Transfer): In the transfer test, the GP group will develop better student maps than the SP and NP groups, and the SP group will develop better student maps than the NP group.

**Hypothesis 2** (Levels of Reflection): The GP group will show higher levels of reflection than SP group.

Participants and Procedures

Participants were 33 students from two local secondary schools (ages ranged from 13 to 15) who took part in the experiments on a voluntary basis for two two-hour sessions within one week: main study and transfer study. They were randomly assigned to one of the three conditions. Eventually, 29 students (76%), 20 female (69%) and 9 male (31%) completed all the activities, resulting in the following division over the three conditions: no prompts (NP) condition as control group: n = 10, specific prompts (SP) condition: n = 10 and generic prompts (GP) condition: n = 9.

The domain in the main study was the elementary economics topics of demand and supply. Economics is both a theoretical and applied domain, seldom studied in class by secondary school students and seldom adopted as the domain in ILE research (e.g. Nichols, 1993). The domain materials were provided to participants before the sessions.
All the three groups completed the main study on economics which was divided into four phases which lasted about two hours: pre-test phase, learning phase, teaching phase and post-test phase. In the pre-test phase, all participants fulfilled the domain knowledge pre-test in supply and demand. Participants received more specific instructions from researchers after the pre-test. In the learning phase, participants were allowed to study the text and diagrams in the materials. They were encouraged to take notes, and were told that their notes and the text would be available during the teaching phase. During the teaching phases, participants were working with the agent tutee system to teach what they have learnt in the learning phase. Both the SP and GP groups were required to write down their reflection notes in the dialog window to respond to the agent prompts to proceed with their teaching activities. Subsequently, all participants completed the post-test domain knowledge assessment which was the same as the pre-test knowledge assessment.

One week later, a transfer test was conducted on the same three groups of students on river ecosystems in ecology (Leelawong, 2005). Transfer is the ability to generalize from the familiar to the less familiar, for example, it is the ability to use problem-solving knowledge learnt in the context or domain of one task in another context or different task (Bransford, Brown, and Cocking 2000). This transfer test was intended to investigate whether students had experienced a transformative change to apply what they have learnt in the domain of economy to understand and solve problems in the domain of ecology. All of them worked with the basic agent tutee without prompts. They studied the materials in river ecosystems first and then began to teach their agent tutee. The river ecosystems include common concepts such as fish, plants, bacteria, dissolved oxygen, carbon dioxide and algae. Each group was asked to teach Betty about river ecosystems to maintain the balance of the natural system.

**Measures**

There were two categories of measures for the two hypotheses: 1) Learning Outcomes, and 2) Response Statements to Agent Prompts.

**Learning Outcomes**

Three measures were used to study students’ learning outcomes as a result of the interventions: Improvements in students’ scores from pre-test to the post-test; quality of the students’ concept maps as compared to expert maps; participant’s ability to transfer.

**Pre- and Post- Test Scores:** The pre- and post- test administered to the students contains 14 questions that include both multiple-choice and true-or-false questions. Question 1, 2, 3, 10 are multi-choice questions which are supposed to check students’ factual knowledge about the domain. Question 4, 5, 6, 7, 8, 9 and 11 are multi-choice questions to check students’ reasoning capacity in determining causal relationships among economic concepts about demand and supply (when something goes up, something else goes down, etc.). Question 12, 13 and 14 are true/false questions to let students make judgments using economic knowledge in a market context of the real world. As to the scoring of these 14 questions, one point is assigned to one question when students marked the correct choice and zero points, otherwise.

**Quality of Student Maps:** We developed a systematic procedure which is adapted from a coding scheme developed by Leelawong (2005) to evaluate the concepts and propositions in student maps which were grouped into two categories: Expert and Relevant. Expert concepts or propositions are those also shown in the expert concept map (Note that students never saw the expert map before). Relevant concepts or propositions are those found in the domain materials or online resources but not a part of the expert map.

**Ability to Transfer:** In the transfer test, students learned the domain of river ecosystems (Leelawong, 2005) and taught the basic version of Betty without the support of agent prompts. Students read the resources and extracted the important concepts and their relations from the text to teach Betty. The measuring procedure of this transfer test is mainly on the assessment of the quality of student maps.
Analysis of Response Statements

An analysis was conducted to analyze participants’ response statements to the agent prompts both qualitatively and quantitatively. The qualitative content of the response statements was examined and quantified to find out at which levels of reflection have occurred in participants’ response statements. In the following subsections, we describe the coding scheme of measuring response statements, the coding process and inter-rater-reliability.

Coding Scheme of Measuring Response Statements

A tripartite coding scheme, adapted by Ortiz (2006) from the categorization scheme proposed by Surbeck, Han, & Moyer (1991), was adopted to analyze the participant’s response statements to the agent prompts. This scheme, which includes categories of reaction, elaboration, and contemplation, adopts the perspective that the nature of the stimulus to reflect will impact the quality of reflection (Moore & Whitfield, 2008), which conforms to what this study seeks to investigate.

In Table 2, the categories of response statements used in this study, tailored from the categorization scheme mentioned above, are provided for comparison and to illustrate and judge participants’ elicited levels of reflection revealed in the learning-by-teaching process.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
<th>Sample Response statements</th>
</tr>
</thead>
</table>
| Reaction/ Reactive Statements | The entry reports facts, feelings, concern of an opinion. Course issues or concepts may be referred with no development or elaboration. | [Respond to how to evaluate the performance of the agent tutee] “Good that you got the answer right, but your explanation isn’t very good. ”  
– Reaction: Report feelings |
| Elaboration/ Elaborative Statements | The entry expands upon descriptive information by explaining feelings, or giving illustrate example of course concepts. The work uses interpretation of course concepts | [Respond to the quest to make explanations to the agent tutee] “Economics fundamentally reflects on supply and demand. Supply is market can offer. It refers to producer supplying of goods.”  
– Elaboration: Interpret course concepts |
| Contemplation/ Contemplative Statements | The entry reference assumption about one’s self or one’s behaviour, indicating a shift in thought or attitude about one’s self. | [Respond to why to teach the agent tutee] “To emphasize the importance of prior knowledge to succeed in future business”  
– Contemplation: Indicate a shift in attitude about oneself |
| Other                       | Statements which are irrelevant or off-topic                               | [Respond to what is the most thing trying to teach the agent tutee] “Music Player. Old and new modern”             |

Coding Process and Inter-Rater Reliability

Two raters were involved in the coding process. To begin, the first rater analyzed a sample of ten response statements to become confident with the coding scheme. Then, the first rater explained the coding scheme, as shown in Table 2, to the second rater. The second rater then analyzed the first ten response statements as well. The consensus was achieved between the two raters to differentiate among the three categories: reaction, elaboration and contemplation.

After that, the two raters analyzed the remaining response statements independently. Cohen’s kappa was used to estimate the level of agreement between the two raters, taking the agreement occurring by chance into account. Calculated with Kohen’s Kappa, agreement between the two raters appeared to be good, $\kappa = 0.653$ (A Cohen’s kappa value between 0-0.4 is considered poor, 0.4-0.6 fair, 0.6-0.75 good, and 0.75 outstanding.).
Results

Learning Outcomes

Pre- and Post- Test Scores

Figure 4 shows the result of pre- and post-test scores. A one-way analysis of variance (ANOVA) result indicated that the three conditions had no significant difference as to students’ prior knowledge when almost all students are domain novices and achieve low scores in the pretest ($F(2, 26) = .219, p = .805$). However, when we look at posttest gains, all the three conditions showed statistical significant increase after the treatment ($F(2, 26) = 3.50, p = .045$).
A one-way analysis of covariance (ANCOVA) test was further used to increase power in the above one-way ANOVA by using pretest scores as covariates. It could control the initial differences in pretest scores to remove "covariate bias" or "selection bias", which weakens internal validity. The ANCOVA test yields a significant effect of the two prompted conditions compared to the non-prompted condition ($F(2, 25) = 19.550$, $p < .05$). However, the pair-wise comparison showed that there was no significant difference ($p = .142$) between the SP and GP groups as to the learning gains from pretest to posttest.

**Map Analysis in Main Study**

The quality of the students’ concept maps in the main study were evaluated to determine whether students’ conceptual understanding of elementary economics varied as they worked on agent tutee with different prompting strategies. Because no significant difference between groups was observed in the pre-test, the subsequent data analysis did not need a covariate. An overall performance comparison among the three groups on relevant concepts, relevant propositions, expert concepts and expert propositions is shown in Figure 5.

ANOVA tests show that there were significant differences among the three groups as to relevant concepts ($F(2, 26) = 29.264$, $p < .05$), relevant propositions ($F(2, 26) = 11.082$, $p < .05$), and expert concepts ($F(2, 26) = 14.358$, $p < .05$). The post-hoc tests of Gabriel’s procedure show that there are significant differences between the NP group and the SP group on relevant concepts ($MD = 2.70$, $p < .05$), relevant propositions ($MD = 2.00$, $p < .05$), and expert concepts ($MD = 2.10$, $p < .05$). There are also significant differences between the NP group and the GP group on relevant concepts ($MD = 2.04$, $p < .05$), relevant propositions ($MD = 1.57$, $p < .05$), expert concepts ($MD = 1.18$, $p < .05$). These results reveal the differences within the three groups. Both of the SP and GP groups showed higher performance than the NP group by working out significantly more relevant concepts, relevant propositions and expert concepts. However, the difference between the GP group and the SP group was not significant.

![Figure 6. Average Number of Relevant/Expert Concepts/Links in Transfer Test](image-url)
Map Analysis in Transfer Test

Figure 6 displays the average number of relevant and expert concepts and propositions in the student maps in ecology. There was no significant difference among the three groups with respect to the number of relevant propositions in the maps. But there were significant differences among the three groups as to relevant concepts ($F(2, 26) = 4.702, p = .018$) and expert concepts ($F(2, 26) = 9.744, p = .001$).

The overall results of the transfer test are reported in Table 3. The post-hoc tests of Gabriel’s procedure on the transfer test show the significant differences between the NP and GP groups on both relevant concepts ($MD = .911, p < .05$) and expert concepts ($MD = 1.256, p = .001$). More importantly, there was also a significant difference between the GP and SP groups on the expert concepts ($MD = .956, p = .010$).

<table>
<thead>
<tr>
<th>Groups</th>
<th>Relevant Concepts</th>
<th>Relevant Propositions</th>
<th>Expert Concepts</th>
<th>Expert Propositions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>NP</td>
<td>2.20</td>
<td>0.632</td>
<td>1.80</td>
<td>.789</td>
</tr>
<tr>
<td>SP</td>
<td>2.60</td>
<td>.516</td>
<td>2.50</td>
<td>.972</td>
</tr>
<tr>
<td>GP</td>
<td>3.11</td>
<td>.782</td>
<td>2.56</td>
<td>.726</td>
</tr>
</tbody>
</table>

*Significant greater than NP, $p < .05$

In sum, the GP group outperformed both the NP and SP groups in the transfer test, while the difference between the NP group and the SP group was not prominent. The results showed that generic prompts have helped students to apply what they learnt from one knowledge domain to another knowledge domain. However, specific prompts did not exhibit transfer effects when time progressed and domain was changed.

Analysis of Response Statements

A detailed analysis of student response statements was conducted to gain a better understanding of the levels of reflection that each experimental group achieved. The prediction was that the students working with generic prompts should produce a higher level than students who worked with specific prompts.

A sample transcript of Participants’ response statement is as follows to demonstrate both SP student and GP student responded to the purposefully designed same question prompts but showed different levels of reflection in their response statements.

[Agent] What is the most important thing you tried to teach me?
[SP Student Response] What is economics? *(Reaction: Report facts)*
[GP Student Response] The universality of these links, causes and implications

*(Contemplation: Indicate a shift in thought).*

[Agent] What is/are the most important thing(s) you learn from me?
[SP Student Response] Demand, supply and price are inter-related and contribute to the idea of economics in total.

*(Elaboration: Provide interpretive analysis of course concepts)*

[GP Student Response] What to do when approaching economics in real life.

*(Contemplation: Indicate a shift in attitude about himself)*

[Agent] Do you learn something from me?
[SP Student Response] Yes, I learn how to organize my thoughts better.

*(Contemplation: Indicate a shift in attitude about himself)*

[GP Student Response] How to teach better and more clearly.

*(Contemplation: Indicate a shift in attitude about himself).*
We further quantified the qualitative analysis results that have appeared in the all the prompted students’ response statements and compared the difference as to the elicited levels of reflection between the two experimental groups.

Table 4 shows a summary of the ANOVA results of analyzing response statement. There were significant difference between the two groups as to reactive statements \( (F (1, 17) = 36.747, p < .05) \) and contemplative statements \( (F (1, 17) = 19.472, p < .05) \). The number of elaborative statements was not significantly different between the two groups. The results exhibited the difference in the levels of reflection between the GP and SP groups. The GP group was more likely to respond with contemplative statements representing a higher level of reflection. Comparatively, the SP group responded more with reactive statements representing a lower level of reflection which means they pay more attention to report issues with no development than the GP group.

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>61.389</td>
<td>1</td>
<td>61.389</td>
<td>36.747</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>28.400</td>
<td>17</td>
<td>1.671</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>89.789</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elaboration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>17.909</td>
<td>1</td>
<td>17.909</td>
<td>4.187</td>
<td>.057</td>
</tr>
<tr>
<td>Within Groups</td>
<td>72.722</td>
<td>17</td>
<td>4.278</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>90.632</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contemplation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>72.047</td>
<td>1</td>
<td>72.047</td>
<td>19.472</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>62.900</td>
<td>17</td>
<td>3.700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>134.947</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.637</td>
<td>1</td>
<td>.637</td>
<td>.872</td>
<td>.424</td>
</tr>
<tr>
<td>Within Groups</td>
<td>16.100</td>
<td>17</td>
<td>.947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16.737</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discussions**

In this study, we explored the inquisitive agent tutee as a learning partner in learning-by-teaching activities. Two types of agent prompts were chosen as the scaffolding approach because secondary school students require a significant amount of support to assume the role of tutor to teach others and handle unfamiliar domains. To focus the study, data analysis concentrated on answering several research questions that involve comparison of the use of an agent prompts and their effects on student’s reflection and learning.

Overall, the preliminary learning outcome results showed that the agent prompts could have the potential to add value and encourage student in reflection and achieving a better learning outcome because the two prompted groups performed better on both pretest-to-posttest test score gains and student map qualities than non-prompted group in this short-term treatment. Generic prompts eliciting double-loop reflection exhibited the effect of cross-domain transfer, while specific prompts eliciting single-loop reflection did not, because the GP group performed better in the transfer test than SP and NP students and no difference was found between the SP and NP groups. However, the differences between generic prompts and specific prompts on learning outcomes in the main study were not clear when the two types of interventions were evaluated.

Participants generated response statements representing a higher level of reflection more frequently when they received generic prompts than specific prompts. Based on the analysis of response statements, we find that generic prompts could promote more contemplative reflection and specific prompts elicited more reactive reflection.

The following sections discuss the comparison between generic prompts and specific prompts on learning outcomes and relationship between agent prompts and reflection in the context of the relevant research literature.
Generic Prompts and Specific Prompts on Learning Outcomes

Reflection can play an important role in improving learning performance (Chen, Kinshuk, Wei, Liu, 2011). Generally, the study does not completely confirm the hypothesis that generic prompts are more effective than specific prompts in terms of learning outcomes. In most cases of this study, the differences between the GP group and SP group were not statistically significant. However, one should note that the GP group outperformed significantly than both NP and SP students in the transfer test when all students used the basic version of agent tutee with prompts removed. It indicated that generic prompts brought more transfer effects to students as time progressed and domain was changed.

Such a result could be explained from the perspective of double loop learning theory. The differentiation of the two types of prompts is respectively based on the single-loop learning to result in an incremental change and double-loop learning to lead to a transformative change, the kind that brings about ultimate changes in an individual or organization (Argyris and Schön, 1978; Akbar, 2003; Moore, Fowler, Watson, 2007). Single-loop learning is to correct an action to solve or avoid a mistake. This type of learning solves problems but ignores the question of why the problem arose in the first place, which has been referred to as “adaptive learning” by Senge (1992). Meanwhile, double-loop learning is to correct the underlying causes behind the problematic action. This type of learning focuses on transformational change that changes the status quo, which has been referred to as “generative learning” by Senge (1992). As Senge (1992) notes, unlike adaptive learning, generative learning requires a new mindset and the capacity to create new visions for future realities. Essentially, generative learning builds in a redesign process based on optimal problem solving. This is in contrast to adaptive learning or single-loop learning where the focus is on solving current problems without examining the root causes of the problem or the learning behaviors that underpin the problem-solving process.

A similar study comparing generic prompts and specific prompts within an Intelligent Tutoring System for legal arguments (Aleven, Pinkwart, Ashley and Lynch, 2006) has a quite similar result. Generic and specific prompts could be both helpful for students to learn in the main study while the difference between the two prompted conditions is not prominent. As Aleven, Pinkwart, Ashley and Lynch (2006) have argued, generic prompts may be useful because they draw the students’ attention to particular passages in the transcript, without restricting students to a small set of inferences. They leave individual student more latitude in discovering deficits in his or her own knowledge. Comparatively, specific prompts may be helpful because they lead students to useful inferences and possible identification of gaps in their understanding (although the latter function is less certain in an ill-defined domain such as the current). They might provide more help in leading students towards specific deficits and possible ways of addressing them. Aleven, Pinkwart, Ashley and Lynch (2006) also note that specific prompts may be harmful in that they are likely to focus students’ attention on a narrower set of inferences than they would otherwise have attended to.

These relevant arguments and findings can be helpful to explain why the GP group can outperform the SP group in the transfer test of this study. Since specific prompts eliciting single-loop reflection only focus on solving current problems, the students in the SP group are not likely to make broader inferences by themselves without the help of specific prompts in the transfer test. Meanwhile, generic prompts eliciting double-loop reflection could bring transformative change for future learning contexts. Thus, the students in the GP group are more likely to make broader inferences by themselves even without the generic prompts.

Agent Prompts and Reflection

The analysis of student’s response statements showed that the GP group was more likely to respond with contemplative statements representing a higher level of reflection. Comparatively, the SP group responded more with reactive statements representing a lower level of reflection which means they paid more attention to the task-specific aspects than the GP group.

Such a result could be linked to the context of reflective practice of pre-service teachers. Mezirow (1981) and Bolton (2001) have emphasized that reflective practice is learning and developmental process enacted through the examination of one’s own practice, including experiences, thoughts, feelings, actions and knowledge. In other words, for pre-service teachers to fully engage in reflection they need to have a clear view of their own philosophy and be
prompted to consider how their beliefs, experiences and knowledge have shaped the theories they apply to teaching and learning. Working from awareness and understanding of their own metacognitive strategies and beliefs, student teachers can then be challenged to consider a wider range of discourses, an aspect that Phelan (2001) argues is vital in recognizing the discourses that shape and often restrict thinking. From our study, it could be possible to conclude that the focus for reflection on metacognitive strategies and beliefs could shape or restrict students’ reflection during the reflective practices.

A similar study is conducted by Andre (1979) who investigates the effects of asking students questions at different levels of cognitive complexity during learning. It is shown that higher level questions can have facilitative effects on both reproductive and productive knowledge, but the conditions under which such facilitation occurs are not well understood. A more recent study by Chen et al. (2009) finds that the main factor affecting reflection levels is high level prompts followed by high quality observation that has a moderating effect on learners' reflection levels.

**Conclusion**

Overall, the results of the study suggest that the use of agent tutee as an active and inquisitive learning partner to ask meaningful questions could be some helpful to students learn by reflective teaching. The preliminary quantitative and qualitative results suggest that the use of an agent prompts as computer-based scaffolds, could be useful in particular ways, such as improving students’ learning outcomes and eliciting higher levels of reflection.

The study suggests that generic prompts could be more beneficial in fostering double-loop learning. There remains significant support challenges that must be overcome before a system similar to that used in this study, with agent prompts, can be incorporated into regular classroom use. Generally, the goal of supporting students needs to shift from a notion of leading to one of facilitating and enabling. This means designing intelligent tutee systems that are not necessarily recipients of information but, rather, they are designed to promote reflection. Instead of attempting to create an agent tutee that will always know the correct answer, designers need to design agent tutee systems that encourage students to do both reflection-in-action and reflection-on-action (Schön, 1987) and pay attention to double-loop reflection (Argyris and Schön, 1974, 1978, 1996), i.e., thinking out of the box.

**Acknowledgements**

We thank Dr. Gautam Biswas in Vanderbilt for providing the Teachable Agent software.

**References**


Lai, G. (2008). Examining the effects of selected computer-based scaffolds on preservice teachers’ levels of reflection as evidenced in their online journal writing. PhD Thesis, Georgia State University, Atlanta, GA.


Utilizing a Collaborative Cross Number Puzzle Game to Develop the Computing Ability of Addition and Subtraction

Yen-Hua Chen¹, Chee-Kit Looi², Chiu-Pin Lin³, Yin-Juan Shao² and Tak-Wai Chan⁴

¹Department of Computer Science and Information Engineering, National Central University, Taiwan // ²Learning Sciences Lab, National Institute of Education, 1 Nanyang Walk, Singapore 637616 // ³Graduate Institute of eLearning Technology, National Hsinchu University of Education, Taiwan // ⁴Graduate Institute of Network Learning Technology, National Central University, Taiwan // oliver@cl.ncu.edu.tw // cheekit.looi@nie.edu.sg // chiupin.lin@gmail.com // yj_shao@hotmail.com // chan@cl.ncu.edu.tw

ABSTRACT

While addition and subtraction is a key mathematical skill for young children, a typical activity for them in classrooms involves doing repetitive arithmetic calculation exercises. In this study, we explore a collaborative way for students to learn these skills in a technology-enabled way with wireless computers. Two classes, comprising a total of 52 students in Grade 4 (ages 10 or 11) participated in the study. They used the Group Scribbles software to run an adapted version of the “Cross Number Puzzle” that was designed with the “feedback” mechanism to assist students’ problem solving. In one class, students played the game individually and in the other class, students played the game collaboratively. The low-ability students in the collaborative class were found to have made the most significant progress in arithmetic skills through playing this game. Three dominant interactive collaboration patterns, one contributing to productive interactions and two to less productive interactions, were also identified in the students’ collaboration.

Keywords

Computer-supported collaborative learning, feedback mechanism, Cross Number Puzzle, computing ability

Introduction

Computational skills in addition and subtraction are an important component in any mathematics curriculum for young children. In seeking to design more interesting learning experiences for children to learn these skills, we designed and implemented a learning activity which is an adapted version of the “Cross Number Puzzle” on a technology platform. The aim of this game is to promote the flexible application of addition and subtraction skills, and to enhance children’s capacity to build up their arithmetic skills progressively. We conducted a study on two classes of students to explore different collaborative learning patterns that involved students working together on solving arithmetic problems. We also examined differences between individual learning and collaborative learning.

A variety of educators have classified operating addition and subtraction story problems into four problem types: change, combine, compare and equalizer (Carpenter, Hiebert, and Moser, 1981; Fuson, 1992; Gustein and Romberg, 1995). English (1998) points out that change and combine are easier while take-away and compare are more difficult challenges for elementary school students. In an arithmetic equation, any of the three numbers could be the unknown number. We adopted this widely used method in our study. Fuson (1992) defines these three types of “change” (placeholder) as: Missing End, Missing Change, and Missing Start. Van de Walle (2001) also classifies these three types of change, combine, compare and equalizer (Carpenter, Hiebert, and Moser, 1981; Fuson, 1992; Gustein and Romberg, 1995). English (1998) points out that change and combine are easier while take-away and compare are more difficult challenges for elementary school students. In an arithmetic equation, any of the three numbers could be the unknown number. We adopted this widely used method in our study. Fuson (1992) defines these three types of “change” (placeholder) as: Missing End, Missing Change, and Missing Start. Van de Walle (2001) also classifies these three types of

With these three levels in Table 1 in mind, we designed our system to incorporate five stages of problem posing to the students (Table 2):

In stage 1, the student is required to derive the answer of an arithmetic expression (result number unknown), inculcating the skills of basic addition and subtraction, for example: 3 ± 2 = □.

In stage 2, the arithmetic operator is removed. Students were required to understand the concept of arithmetic operator, for example: 3 □ 2 = 5.

In stage 3, the change amount is removed, for example: 3 ± □ = 5.
In stage 4, the initial amount is removed, for example: $\Box \pm 3 = 5$.
In stage 5, both the initial and the change numbers are removed. It is more difficult as the sentence includes two variables: change amount and initial amount unknown. For example: $\Box \pm \Box = 5$.

**Table 1. Three levels of “Change” types in story problems**

<table>
<thead>
<tr>
<th>Change Types</th>
<th>Join (Add to)</th>
<th>Separate (Take from)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result number unknown</td>
<td>John has 5 cookies; Marry gave 5 cookies. How many does John have now? Standard sentence: $A + B = \Box$</td>
<td>John has 8 cookies; he gave Marry 5 cookies. How many does John have now? Standard sentence: $A - B = \Box$</td>
</tr>
<tr>
<td>Change number unknown</td>
<td>John has 5 cookies; Marry gave some cookies, and now John has 8. How many did John get from Marry? Standard sentence: $A + \Box = B$</td>
<td>John has 8 cookies; he gave Marry some cookies, and now John has 5. How many did John give to Marry? Standard sentence: $A - \Box = B$</td>
</tr>
<tr>
<td>Initial number unknown</td>
<td>John has some cookies; Marry gave 5 cookies, and now John has 8. How many cookies did John have before Marry gave him some? Standard sentence: $\Box + A = B$</td>
<td>John has some cookies; he gave 5 cookies to Marry, and now John has 3. How many cookies did John have before he gave some to Marry? Standard sentence: $\Box - A = B$</td>
</tr>
</tbody>
</table>

**Table 2. Level of difficulty in the design**

<table>
<thead>
<tr>
<th>Level of difficulty</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Result number unknown — basic skill practice</td>
<td>$A \pm B = \Box$</td>
</tr>
<tr>
<td>Level 2</td>
<td>Remove operator — between basic skill practice and comprehension application</td>
<td>$A \Box B = C$</td>
</tr>
<tr>
<td>Level 3</td>
<td>Change number unknown add-to or subtraction — comprehension application</td>
<td>$A \pm \Box = B$</td>
</tr>
<tr>
<td>Level 4</td>
<td>Initial number unknown add-to and subtraction — comprehension application</td>
<td>$\Box \pm A = B$</td>
</tr>
<tr>
<td>Level 5</td>
<td>Change number unknown and Initial number unknown, addend or summand type — the most difficult level</td>
<td>$\Box \pm \Box = A$</td>
</tr>
</tbody>
</table>

**Design of the “Feedback” system**

A ‘feedback’ mechanism was introduced to the game design in this study. Feedback is considered to have strong impact on the learning process and result (Kulhavy and Stock, 1989; Bangert-Drowns, Kulick, Kulik and Morgan, 1991; Balacheff and Kaput 1996). Appropriate feedback can lead learners to focus on key elements of learning. They can always adjust their learning strategies to try to close the gap between their actual performance and the learning goals. They reflect on their learning by a self-monitoring feedback loop. Hence, they can change their learning strategies in the follow-up learning and seek a better way of learning (Alexander and Shin, 2000). The learner understands the gap or difference between her own concept and the target concept, and the perceived difference can provide information for her to amend her ideas and knowledge, thus enacting a process of signal feedback.

Schmidt (1991) proposes that feedback is the result of a series of actions. It represents the personal response or reaction to the information they received. The feedback itself is a problem solving process that checks the performance of action to improve a person or a group. In technology-enabled learning, feedback is typically provided as messages shown to students after their responses (Cohen, 1985). Siedentop (1991) points out that feedback can promote the interaction between the teacher and the learners. Teachers can provide feedback to students in terms of their actions and performance, which enable them to know or to amend their understandings and may boost their enthusiasm for learning (Keh, 1992).

There are three forms of feedback: immediate feedback, summary feedback and compromise feedback (Schmidt and Wrisberg, 2000). Collins, Carnine, and Gersten (1987) point out three levels of feedback messages: (1) little feedback: just show the answer is right or wrong; (2) basic feedback: if answer is not correct then show right answer; and (3) descriptive feedback: give some hints to learner to help her to obtain the right answer. Descriptive feedback can increase the motivation for learners to try or solve new tasks and new problems. The feedback mechanism provided by software systems mainly involves five levels such as: no feedback, knowledge of response, knowledge
of correct response, answer until correct, and elaboration feedback, as summarized by Sales (1998). In our research, the feedback mechanism is designed as follows (Figure 1):

**Flow chart**

<table>
<thead>
<tr>
<th>Knowledge of response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer until correct</td>
</tr>
<tr>
<td>Knowledge of correct response</td>
</tr>
<tr>
<td>Answer until correct, AUC</td>
</tr>
<tr>
<td>Elaboration feedback, EF</td>
</tr>
</tbody>
</table>

**Description**

- Feedback to the student regardless of whether her response is correct or incorrect
- Set the upper limit of number of incorrect attempts before providing a new clue
- Point out what should be amended
- Set the upper limit of number of incorrect attempts before providing a new clue
- Provide the detailed answer, explain and elaborate the answer

*Figure 1. Five-level feedback mechanism*

Progressive feedback is provided in a step by step manner in this five-level mechanism. In the ‘knowledge of response’, the student get a response whatever she replies correctly. Then a constraint on the upper limit of incorrect attempts is invoked to allow some trial and error. If the student fails after a number of trials, a clue or a suggestion would be offered to her for another round of trial and error. In the ‘elaboration feedback’, the correct answer together with detailed explanations will be provided.

**Methodology**

Fifty two students in Grade 4 (ages 10 or 11) participated in our study. They learned some basic addition and subtraction since Grade 1 but they needed to connect this to the new concepts and skills required for Grade 4 mathematics. In this research we explored the effects of “Cross number puzzle” game applied in learning, which was designed to provide the feedback mechanism. We had two experimental classes: students in Class A played the “Cross number puzzle” game in small groups, and students in Class B played the game individually. All students were grouped according to their average scores of the previous three tests in this term. Using percentile ranking, those students with the percentile rank of score over 73% were classified as high-math achievers; those with percentile rank of score between 27% and 72% were classified as medium-math achievers, and those with percentile below 26% were classified as low math achievement. Students in class A were divided into homogeneous groups with three per group. 6 students in the high-achiever group form two groups. Three medium-achiever groups comprised 9 students and another three low-achiever groups comprised 9 students.

We utilized Group Scribbles (GS) as the platform for the game, and conducted analysis of the collaborative work within these groups. GS is a computer-supported collaborative learning system developed by SRI International to conduct small-group collaborative concept mapping activities (Chaudhury, et al, 2006; Looi, Chen and Ng, 2010). Each student has a Tablet PC which has a screen divided into upper and lower frames (Figure 2). The lower frame is for individual cognition, that is, the student sketches or types her answer individually. The upper frame is a shared space (public board) in which the students show all of their individual answers, and work together as a group. They can even check the work from other groups by clicking the button on the top right corner (see Figure 2). The teacher can monitor their process of learning and provide appropriate guidance.
Figure 2 shows the interface of “Cross Number Puzzle” in GS. We designed questions ranging from the easy to the difficult in terms of the five levels of difficulty. When the students complete the calculation, they can fill in the answer box and press OK button under the question area to submit. If the answer is correct, there will be a brief description of the key points. If the answer is wrong, the system will execute a step-by-step tip based on the number of errors from the user inputs (Figure 8, a popup box with tips shown in individual area). The action repeats until the maximum number of errors reaches the upper limit. Then the system will show the correct answer and the methods of problem-solving. Four different types of questions were shown below (Figure 3 to Figure 6) in the “Cross Number Puzzle”.

Figure 3. Question type 1

Figure 4. Question type 2

Figure 5. Question type 3

Figure 6. Question type 4

Figure 7. Interface of individual and public work

Figure 8. Tips

Figure 9. Calculation process
Figure 7 shows a screenshot of one student’s calculation process. Questions appeared in the private board and the individuals could paste their sketches to the public board. Figure 9 provides the screenshot of the calculation process in a four-member group, in which each individual pasted his/her sketches using a different color.

The study for both classes lasted for four weeks. In the first week, a session for 30-minute pre-test and 20-minute training was executed. Students were asked to be familiarized with GS and the operation of the game with simple exercises. In the second week, the game was played in one lesson lasting for 60 minutes, followed by a 30-minute post-test and a 20-minute questionnaire in the third week. In the fourth week we interviewed the teachers and students. A pre-activity and three learning activities were included in this study.

In the game playing session, three tasks were designed and implemented in collaborative group (Class A) and individual group (Class B) separately. Students were asked to fill in the operator in an arithmetic equation in Activity 1. Activity 2 is about filling in the unknown number while in Activity 3, students were asked to estimate using trial-and-error methods to solve the problem. The only difference between these two classes is that students in Class B played the game individually but students in Class A played it collaboratively. Figure 10 below gives an example of a game screenshot of a collaborative group with four members. Group members could use their private boards for sketches and confirmatory calculations. Then they could post their sketches or results to the public board. Figure 11 shows an example of a student’s work from a student who worked individually. We can inspect this student’s solution path to identify and understand his working strategy.

Findings

We analyzed students’ scores in pre and post-tests, collected questionnaires, did video-recording of their activities in classes and tracked their screens in the process of game playing. The pre-test and post-test have the same questions but they are ordered differently. We also verify the item discrimination index and had the questions validated by a experienced mathematics teacher.

Results of the assessment of computing ability

We administered pre-tests and post-tests and performed independent sample t-test of two classes on their results.

Table 3 shows the pre and post test results of Class A. Students in Class A have higher average scores in the post-test. Their average score was increased by 13.00, from 50.29 in pre-test to 63.29 in post-test (p=0.002<0.01). This indicates students in Class A made greater progress than those students in Class B through playing the game collaboratively. Further observation of these collaborative groups suggests that the low math-achiever students made the most significant progress, which can be easily gathered from the following table.
Table 3. T-test of pre and post tests in collaborative group

<table>
<thead>
<tr>
<th>Number of Participants (N=24)</th>
<th>t-test of collaborative group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests</td>
<td>Number of Participant</td>
</tr>
<tr>
<td>Pre-test of high-achievers</td>
<td>6</td>
</tr>
<tr>
<td>Post-test of high-achievers</td>
<td>48.34</td>
</tr>
<tr>
<td>Pre-test of medium achievers</td>
<td>9</td>
</tr>
<tr>
<td>Post-test of medium-achievers</td>
<td>40.01</td>
</tr>
<tr>
<td>Pre-test of low-achievers</td>
<td>6</td>
</tr>
<tr>
<td>Post-test of low-achievers</td>
<td>16.67</td>
</tr>
</tbody>
</table>

The low-achiever groups in Class A were found to have the highest increase in post-test scores at a high level of significance (P=0.001). This indicates that low-achievers of these collaborative groups derived the most benefits in this study.

Table 4 showed further analyses conducted on three different types of test questions on “addition and subtraction”. Students had better scores in all three types of questions in the post-test. But the low-achiever groups achieved significantly highest improvement in questions of “basic computing”, “unknown constant” and “Cross Number Puzzle” with the increase of average score 9.63, 7.38 and 6.32 respectively. This suggests that these low-achievers benefited the most from the “Cross Number Puzzle” in improving their basic arithmetic skills.

Table 4. Low achieved students’ progress in Pre and Post tests

<table>
<thead>
<tr>
<th>Low achievers in Class A (N=9)</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Average increased scores</th>
<th>Ratio of progress in different questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic computing skills</td>
<td>Score of question 1 to 5 (33.33)</td>
<td>18.52</td>
<td>28.15</td>
<td>9.63</td>
</tr>
<tr>
<td>Unknown constant</td>
<td>Score of question 6 to 12 (46.67)</td>
<td>8.90</td>
<td>16.28</td>
<td>7.38</td>
</tr>
<tr>
<td>Cross number puzzle</td>
<td>Score of question 13 to 15 (20.00)</td>
<td>3.14</td>
<td>9.46</td>
<td>6.32</td>
</tr>
</tbody>
</table>

On the other hand, the average score of Class B (the individual group) is 4.17 higher in the post-test (57.21) than in the pre-test (53.04) at a significant level of .026 (p<0.05). This indicates that learners in Class B also made progress through playing the game.

To gain further insight about the differences of improvement between the individual group and the collaborative group, regression analysis was done to judge the relationship between the scores of pretest (as the independent variable) and the score of post-test (as the dependent variable) within both classes. The results of F-test for pre-test (F = 2.487, p = .121, p > .05) shows that the individual group (Class B) and the collaborative group (Class A) can be regarded homogeneous.

However, Table 5 below provides the result of F-test (F = 4.479, p = .039, p <.05), which is significant at the p<.05 level. The striking result to emerge from the data is that the collaborative group had much greater improvement than the individual group in this study although they played the same game.

Table 5. ANOVA for individual group and collaborative group

<table>
<thead>
<tr>
<th>Analysis of variance for Class A and Class B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
</tr>
<tr>
<td>Inter-group</td>
</tr>
<tr>
<td>In-group</td>
</tr>
</tbody>
</table>
Collaboration

Questionnaire results illustrated 85% students tried to do cooperation and discussion before they submitted the answer when they played the “Cross number puzzle” game. There was one high-achiever student who did not discuss with others when he did his calculations. He explained in the follow-up interview that he was quite confident and only shared his results with others when he completed all his calculations. 87.5% students claimed that it was much easier to complete the calculations with collaboration than doing them individually. Those students without confidence in mathematics found it easier to share their own ideas with others and complete the calculations together. All students agreed that they derived benefits from discussion with other classmates.

“Tips” usage in Class A and Class B

As we mentioned before, students in Class A play the game collaboratively in groups while students in Class B completed the game individually. We can easily conclude from Table 5 below that feedback in the form of “Tips” was much more frequently used in Class B than in Class A. It suggests that when a student encounters problems and difficulties and in the situation where there is no help from peers, he or she would search help from the “feedback” system. On the other hand, students in Class A would discuss their strategies to solve the problem within a group first, allocating cooperative work among group members. They only referred to the “feedback” system when every student in the group was uncertain on how to proceed. They used the “tips” less often then students in Class B. Students in Class B seldom initiatively asked for help from their classmates but only referred to the tips when they encountered difficulties. However either in Class A or Class B, high-achiever students seemed to have used the “tips” far less than low-achiever students. Low achieving students relied more on “tips”.

<table>
<thead>
<tr>
<th>Number of use in different group</th>
<th>Class A (N=24)</th>
<th>Class B (N=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-achiever</td>
<td>0.54</td>
<td>0.84</td>
</tr>
<tr>
<td>Medium-achiever</td>
<td>0.71</td>
<td>1.31</td>
</tr>
<tr>
<td>Low-achiever</td>
<td>1.25</td>
<td>2.09</td>
</tr>
<tr>
<td>Average usage</td>
<td>0.86</td>
<td>1.40</td>
</tr>
</tbody>
</table>

We analyze further students’ collaborative activities (in Class A) in the process of game playing which included three tasks to be completed in the game.

Pre-activity: Result number unknown

Before formal learning activities, a pre-activity was run with two objectives: familiarization with the GS system operation and warming-up the student work in doing arithmetic calculations. The question model is A ±B =□ which is the level one difficulty. 23 of 24 students in collaborative groups finished this activity.

Learning Activity 1: Remove the operator

Four different patterns of collaborative problem solving were found in their activities of “remove the operator”: whole-group-deciding, two-member-deciding, leader-deciding and individual deciding. Group 6 had the decision made by all group members. Three groups, Group 1, Group 3 and Group 7 decided the answer on an individual basis. Two groups followed the two-member deciding pattern and the other rest two groups the leader-deciding pattern. The following figures (Figure 12 to Figure 15) shows different layout of the game in different collaborative methods. For example, in figure 11, three students in group 6 (one student in one color of “+”) posted their answer as 4777 +++ 4611 +++ 1799 = 11154, six “+” and one “=”. All these three students could perform the addition operation correctly. Therefore we could judge that this group’s answer was decided by the whole group. In Figure 14, there is only one answer being pasted, and when we referred to the video recording, we found this group was clearly one of leader-deciding.
Learning Activity 2: Fill in the figure in the formula sentence

To enable learners get the unknown number in the puzzle by observing, calculating those given numbers and estimating the result, for example, \( A \pm \square = B \) & \( \square \pm A = B \), tasks division and coordination were necessary in one group. From the procedural layouts of the game on the screen we got some insights of methods of students’ collaboration and their strategies to complete the calculation. The results were shown in Table 7.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Description</th>
<th>Group</th>
<th>Ratio of different method used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual calculation</td>
<td>Group members did the calculation by themselves individually.</td>
<td>G1, G2</td>
<td>25.0%</td>
</tr>
<tr>
<td>Comparison</td>
<td>Started from different thread and compare each other’s result at the intersection</td>
<td>G3, G5</td>
<td>25.0%</td>
</tr>
<tr>
<td>Relay</td>
<td>One finish one section and another take over to continue calculating</td>
<td>G4</td>
<td>12.5%</td>
</tr>
<tr>
<td>Assisted calculation</td>
<td>One of the group members is in charge of all calculation and other members checking his/her calculating process</td>
<td>G6</td>
<td>12.5%</td>
</tr>
<tr>
<td>Through-out calculation</td>
<td>Some members calculate from the beginning to the end and other members calculate from the end to the beginning then they compare at the intersection.</td>
<td>G7, G8</td>
<td>25.0%</td>
</tr>
</tbody>
</table>

Learning Activity 3: Fill in Multi-unknown

This most difficult task (Level 5), students will fit in multiple unknowns equation like \( \square + \square = C \) or \( \square - \square = C \). After analysis student’s working path, trace group’s problem solving strategies. The most frequently used collaboration strategy is cross calculation (about 50%). Table 8 shows the results.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Description</th>
<th>Group</th>
<th>Ratio of different method used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual calculation</td>
<td>Group members did the calculation by themselves individually.</td>
<td>G1</td>
<td>12.5%</td>
</tr>
<tr>
<td>Cross calculation</td>
<td>Started from different paths and compare each other’s result at the intersection</td>
<td>G2, G3, G4, G5</td>
<td>50.0%</td>
</tr>
<tr>
<td>Reverse calculation</td>
<td>Some members calculated following vertical or horizontal paths, other members calculated from the result to get the answers.</td>
<td>G6, G7, G8</td>
<td>37.5%</td>
</tr>
</tbody>
</table>

Interactive collaboration patterns

By analyzing the video recordings of the students’ activities together with the recording of computer screens in Class A, different interactive patterns in eight groups were identified. We will illustrate these interaction patterns in the
following diagrams. The double-arrow ↔ represents the full dual interaction between two learners. For example, if student S1 communicates much with student S2, we mark it with S1 ↔ S2. The arrow → represents strong one-way interaction and the dashed arrow → represents weaker one-way interaction. For example, if student S5 tries to discuss with student S6, but S6 does not respond, we mark it with S5→S6. If student S4 talks often to student S3, but student S3 does not respond, we mark it with S4→S3.

Milson (1973) identified seven frequently occurring interactive patterns within small learning groups, namely:

1. Unresponsive: The leader communicates with other members but they do not respond;
2. Unsocial: There is no interaction among the group members;
3. Dominant leader: The leader communicates with group members who revere and follow the leader;
4. Tete-a-tete: Every member interacts with his/her neighbors, which may lead to emergent sub-groups;
5. Fragmented, cliquish: Fragmented interactions within individual sub-groups but no interaction at the full-group level;
6. Stilted: Each and every member interacts with his/her neighbors; the interactions have yet to reach an ideal status albeit some individuals have opportunities to interact with each others;
7. Ideal: All the group members actively interact with each others, and there are multiple communicative paths.

Three of Milson’s interactive patterns were identified in our study.

The ideal interactions

As shown in Figure 16, Group 3, Group 4, Group 5, Group 6 and Group 8 had full dual communication when they had discussions. Before they started to do the calculations, they had a discussion and allocated work amongst the group members. Everyone in the group was engaged in the discussion and helped each other solve problems, and this alleviates any difficulties encountered by the students in system operations or calculations. Because of the successful implementation of the division of labor and good communication, these groups tended to have good performance of learning activities.

In Figure 16, the percent of work completed means total progress in all activities, for example, Group 3 had finished 100% questions in pre-activity, learning activities 2 and 3, but completed 93.75% for learning activity 1, so this
group’s progress is 98.4% \((1+1+1+0.9375)/4\). Group 8 finished all problems in all activities and got 100% progress.

The fragmented interactions

Group 2 and Group 7 manifested this interactive pattern, as illustrated in Figure 17. Analysis of the video recordings indicated that they did not have explicit cooperation at the beginning. Some members in these groups started individual calculations. They only discussed when they encountered difficulties and problems. Once the problem was solved, they went back to individual work. Some members of the group had weak communication or one-way communication with other members. At times a member of the group appeared isolated. Errors and omissions occurred due to their unsuccessful communication resulting in unproductive sharing.

**Figure 17. Fragmented interactions in Group 2 and Group 7**

The unresponsive interactions

The interaction in Group 1 was unresponsive as shown in Figure 18. Video analysis shows that at the beginning one of the members was quite active. He tried to allocate the task to another group member. However both group members paid attention only to their individual calculation and neglected each other suggestions. This learner obtained little useful help from his group members and the whole group performed not well. They only finished 66.88% of the work.

**Figure 18. Unresponsive interactive pattern in Group 1**

Teacher’s voices

The students of Class A were motivated and had good sharing in their group tasks. The teacher of Class A shared his experience with us in our interview with him:
“Most students were encouraged to have more discussions in this class. One of the high-achiever groups had conflicts during the discussion because everybody exhibited high confidence and expectation. Another medium-achieving group showed great enthusiasm in collaborative learning with one of them playing the role as a leader.”

“Every student in Class A could get feedback from the system as well as from other members. The consensus achieved in the group made the whole class improve. However in Class B, students had great diversity in their responses. Some students produced good responses when they understood but for some others, they were not sure and thus they talked with their neighbors. And some of them just immersed themselves in individual work and require the teacher to guide them when they encountered difficulties.”

Figure 19 to Figure 22 illustrate some of students’ activities in the class.

**Conclusion and discussion**

This paper investigated the benefits of learning addition and subtraction through the game “Cross Number Puzzle” on Group Scribbles. Our observations and investigations of the two classes who played the game individually and collaboratively respectively showed some interesting differences.

**Effects of collaboration**

The collaborative learning groups (Class A) were found to have made greater progress than individual learning groups (Class B). It suggests that collaborative learning may have enhanced learning effectiveness. From the statistics, we can conclude the low-achiever students benefited the most in this “cross number game”. Collaboration also plays an important role in enhancing learning in Class A with the incorporation of the “feedback system” and collaboration strategies.

**Benefits of feedback system**

In both classes, the low-achiever students accessed the “tips” most often while the high-achiever the least. The individual learning groups in Class B had much higher frequency of access to “Tips”. It indicates that the collaboration among group members in Class A did assist students’ problem-solving. They relied less on the “feedback” system because they could get help from group members. However, in both classes, the low-achiever students had the highest demand for “tips” for help.

**Methods of problem solving**

Students in collaborative learning groups presented four different methods of problem solving in their activities of “removing the operator”: whole-group-deciding, two-member-deciding, leader-deciding and individual deciding. In
the activity of “fill in the figure in the expression,” the students had five methods of calculations: individual calculation, comparison, relay, assisted calculation and through-out calculation. Students also showed four different ways of calculation: free calculation, calculate from the top, calculate from the bottom and calculate from both the top and bottom. They did the calculation in three different collaborative ways: each student calculates the whole thing him/herself; one student started from the top and the other started from the bottom; and they did backwards calculation for checking.

Interactive patterns

Three interactive patterns were found in this study. Among these, the ideal interaction occurred most often. The students all did well in their collaboration. The groups doing fragment and unresponsive interaction were not as interactive as the ideal groups. They had fewer communication and little cooperation.

Future work

Based on these findings in this study, we make the following recommendations for future research.

More experiments to further probe the “Cross Number Puzzle”

Our study has limitations of time and scale. To make the cross number puzzle more applicable, we may need to do more experiments and expand the number of users.

Larger shared screen

A big screen to display the public board for all group members could assist the discussion within a group by providing a focal point of attention. The individual board could still be retained in the screen of the students’ personal Tablet PC for their private cognition.

Adaptive feedback

We only offered phased hints to students in this “Feedback system”. The feedback only includes the general direction of calculation concept and the problem solving process. If system can diagnose and evaluate the individual student’s errors, system can provide each student with the individual corresponding solutions or suggestions to fit his skills.

Incorporating a timer

From our analysis of the frequency of feedback in this study, a timer could be added to the system to record the duration of problem solving by each user. This would also enable the teacher to gauge the time used by the students at each stage of their problem-solving.

References


Effects of Speech-to-Text Recognition Application on Learning Performance in Synchronous Cyber Classrooms

Wu-Yuin Hwang, Rustam Shadiev, Tony C. T. Kuo* and Nian-Shing Chen

Graduate Institute of Network Learning Technology, National Central University, Taiwan // 1Department of Management & Information, National Open University, Taiwan // 2Department of Information Management, National Sun Yat-Sen University, Taiwan // wyhwang@cc.nceu.edu.tw // rshadiev@lst.nceu.edu.tw // tony@mail.nou.edu.tw // nschen@faculty.nsysu.edu.tw

*Corresponding author

ABSTRACT

The aim of this study was to apply Speech-to-Text Recognition (STR) in an effort to improve learning performance in an online synchronous cyber classroom environment. Students’ perceptions and their behavioral intentions toward using STR and the effectiveness of applying STR in synchronous cyber classrooms were also investigated. After the experiment, students from the experimental group perceived that the STR mechanism was easy to use and useful for one-way lectures as well as for individual learning. Most students also expressed that they were highly motivated to use STR as a learning tool in the future. Statistical results showed moderate improvement in the experimental groups’ performance over the control group on homework accomplishments. However, once the students in the experimental group became familiar with the STR-generated texts and used them as learning tools, they significantly outperformed the control group students in post-test results. Interviews with participating students revealed that STR-generated texts were beneficial to learning during and after one-way lectures. Based on our findings, it is recommended that students apply STR to enhance their understanding of teachers’ lectures in an online synchronous cyber classroom. Additionally, we recommend students should take advantages of the text generated by STR both during and after lectures.

Keywords

Speech to text recognition, Synchronous learning, Homework, Note-taking

Introduction

A number of studies have reported the benefits of online synchronous teaching and learning for online courses, although some challenges and limitations still require resolution (Chen, Ko, Kinshuk & Lin, 2005; Hastie, Hung, Chen & Kinshuk, 2010; Wang, Chen & Levy, 2010). One of the most common concerns reported is the presence of poor audio quality due to restricted bandwidth availability and traffic congestion for last mile of Internet access (Chen et al., 2005; Hastie et al., 2010; Wang et al., 2010). According to Chen and Wang (2008) and Kanevsky et al. (2006), students who suffer from bandwidth problems during online synchronous lectures can benefit from reading text streams, which may be synchronously typed on a keyboard or transcribed by Speech-to-Text Recognition (STR) technology. Moreover, Chen and Wang (2008) and Wald (2010) emphasized the pedagogical usefulness of text displayed simultaneously for students during a synchronous lecture, as it facilitates better learning. However, previous research tended to focus on issues related to STR application development and its rate of recognition accuracy improvement, rather than on how it can be applied for improving learning performance (Kanevsky et al., 2006; Wald & Bain, 2008; Way, Kheir & Bevilacqua, 2008). Furthermore, most studies only applied STR in a traditional face-to-face teaching setting but not in an online synchronous teaching and learning environment (Ryba, McVor, Shakir & Paez, 2006; SRS, 2011; Wald, 2010).

This study argues that teaching and learning activities in an online synchronous cyber classroom can be better facilitated by using STR technology. For example, students can follow a teacher’s lecture more easily by reading the STR-generated texts, if the quality of audio degrades during communications; therefore, STR-generated texts can minimize audio communication difficulties and reduce any chance of missing important information. In the interim, STR-generated texts can help students attain a better understanding of a lecture’s meaning, allow for simultaneous note-taking during the lecture, and help students to complete homework after the lecture.

An experiment was conducted with the aim to apply STR for improving learning performance in an online synchronous cyber classroom environment as well as in a situation wherein an individual student completed homework. Students’ perceptions and behavioral intentions toward using STR and the effectiveness of applying STR on learning performance in a synchronous cyber classroom were also investigated. This study addressed three
primary research questions. First, what are the students’ perceptions and behavioral intentions regarding the use of the STR technology in a synchronous learning environment? Second, do the students who use STR-generated texts perform better at accomplishing homework tasks and in post-test evaluations than the students who do not use STR technology? Third, based on interviews, are STR-generated texts beneficial to students?

The remainder of the paper is organized as follows: literature regarding synchronous teaching and learning with STR are reviewed. The theories or models, which are used to analyze the usefulness of the designed STR mechanisms, are discussed. The description of the study’s designed method follows. The results and pedagogical implications of the study are then presented. Finally, a few concluding remarks are given.

**Literature Review**

**Online Synchronous Teaching and Learning**

A number of studies have demonstrated the benefits of online synchronous teaching and learning for online courses (Chen et al., 2005; Hastie et al., 2010; Wang et al., 2010). For example, Hastie et al. (2010) argued that online synchronous teaching and learning allows teachers and students to establish their communications link, create a social presence, and negotiate learning content. They then embark on the ‘real’ synchronous interactive component of the lesson and the teacher gives the students ‘live’ feedback and evaluation. According to Chen et al. (2005), “…in many situations synchronous solutions for instruction can outperform both asynchronous online instruction and traditional face-to-face education.”

However, advantages of online synchronous teaching and learning can be hindered by certain challenges and limitations. Students in the studies of Chen et al. (2005), Hastie et al. (2010), and Wang et al. (2010) occasionally experienced technological challenges during online synchronous learning caused by network traffic congestion. Specifically, the students could not hear the audio clearly.

Chen and Wang (2008) drew attention to text chatting during synchronous teaching and learning by emphasizing that it can accommodate interaction in many important ways. First, text chatting can be used to supplement and complement audio when its quality becomes problematic. Second, text chatting can be used by the teacher or by an advanced student to summarize the major points of verbal exchange to review what has been previously said.

**Technology of STR and its Accuracy**

STR technology translates speech input into text in real-time. Way et al. (2008) described STR as a process. This process involves a teacher speaking into a microphone, wherein the speech is recognized and shown synchronously in the form of text for students to read. The accuracy rate is considered one of the fundamental issues in STR studies. Wald and Bain (2008) claimed that the accuracy rate of STR during a lecture depends on a teacher’s lecture experience, abilities, and familiarity with the lecture material. Way et al. (2008) suggested that STR application training should take place in order to achieve good dictation accuracy. Their study demonstrated that the accuracy rate reached 90 percent after moderate STR training. The accuracy rate of application reached 91 percent after the dictionary of STR was customized with unfamiliar domain-specific terminology.

**STR for Education**

STR has the potential to become a valuable tool in education as it makes teaching accessible and understandable to all students and it improves the quality of education (Wald, 2010). For example, the Speech Recognition in Schools (2011) project helped students to overcome difficulties in reading, writing, or spelling; the project reported significant improvements in some students’ basic reading, writing, and spelling skills with the support of STR. Ryba et al. (2006) examined STR in the university lecture theatre. Their participants reported that the system had potential as an instructional support mechanism; however, a greater accuracy in the system’s recognition of lecture text vocabulary needs to be achieved. Wald (2010) and Wald and Bain (2008) developed STR applications to assist deaf students and non native speakers during lectures. According to this research, the students perceived that text
generated by STR could improve learning as long as it was reasonably accurate (i.e., >85 percent). Most students used STR-generated text as an additional resource to verify and clarify what they heard as well as to take and augment their own notes. The students also believed that lecture transcriptions helped them to better understand the lecture content.

**Method**

This study used the research method, learning design, STR application and training, and experimental research.

**Research Method**

The method of this study was based on three major steps, as shown in Figure 1. The first step included pretesting. The second step involved experimental treatment. The third step incorporated homework assessment and post-testing.

![Flowchart structure for the study](image)

**Learning Design**

This study conducted an experiment with appropriately designed learning activities by using STR technology in an online synchronous lecturing environment as well as in a setting wherein an individual student completed homework.

*Online one-way lectures*

This study used the JoinNet™ application (Wang et al., 2010) to support online learning activities, i.e., online synchronous one-way lectures in a synchronous lecturing environment. The JoinNet™ application provides such tools as a whiteboard, a chat box, and audio and video for synchronous communication purposes (see Figure 2). Additionally, a teacher may upload and present PowerPoint® slides in a synchronous cyber classroom using the whiteboard. This study employed a chat box that displayed STR-generated text to the experimental group. Originally, the teacher planned to use the STR application during online one way lectures and he hoped the application could generate the text in real time. However, Petta and Woloshyn (2001) cautioned about significant delays (up to 50 seconds) in transcript generation by the STR application. Thus, the teacher in this study changed the original and ideal approach and prepared the lecture transcripts beforehand using the STR application in order to reduce delay in transcript generation. Afterwards, the teacher copied STR-pre-generated text and inserted it inside the chat box’s input field during the actual synchronous lectures. Finally, the teacher pressed the Enter key of his keyboard for the STR-generated text to display inside the chat box, so then all students could read it simultaneously.
Individual learning

Individual learning for students included studying the content of previous synchronous lectures and doing homework after online classes. The content of previous synchronous lectures was recorded by the JoinNet™ application and stored online so students could review it. Cooper (2007) argued that homework has an immediate effect on the retention and understanding of the material it covers; thus, students completed homework to practice and master the teaching material. According to the instructional design theory, homework represents skills, knowledge, and even attitudes of students, which are stated in the instructional objectives. Smith and Ragan (2004) recommended testing complex, “high order” knowledge and skills in the real-world context they are actually used, generally with open-ended tasks, such as an essay writing. Therefore, this study employed essay writing for the students’ homework. The first homework session (HW1) related to the “e-Data Protection” synchronous lecture, and the second homework session (HW2) related to the “File Protection” synchronous lecture. The essay writing activity included: 1) summarizing an online one-way lecture; 2) generating and articulating an understanding and opinion about general concepts of the lecture; and 3) elaborating on individual knowledge and experience. Such a writing activity can challenge students to approach, learn, and explain the complexities of the subject matter in new and thought-provoking ways (Smith & Ragan, 2004).

STR and Training

This study employed Windows® Speech Recognition using the Microsoft® Operating System for STR tools. The choice was made based on the availability of this application for the students and teacher participating in the experiment. Way et al. (2008) argued that this application is similar to a variety of commercial and open-source
products in performance, ease-of-use, and it is available at no additional cost. Following general recommendations of Wald and Bain (2008), the teacher started using the STR application 2 months before the experimental course. He trained on the system first, and then applied it in an online synchronous cyber classroom environment after achieving an STR accuracy rate of more than 90 percent. The experimental group was engaged in STR training during the first month of the experiment. These students were asked to use STR after being trained to complete homework and to consequently identify strengths and limitations of the STR application. In addition, STR training could help students participate in a post-study experiment to investigate the effectiveness of applying STR on improving learning performance in a synchronous student-centered learning environment.

Research of the Experiment

The research of the experiment was based on participants and procedures, experimental design, in addition to experimental tools and methods of statistical analysis.

Participants and experimental procedures

A total of 44 undergraduate students enrolled in an Information Security course from two classes were participated in this study. One class with 19 students served as the control group, and the other class with 25 students served as the experimental group. Table 1 shows the participants’ gender and age distributions in two groups. An Information Security course was administered in this study from March to June 2009. A 2-hour, physical, face-to-face or online class took place on a weekly basis. Physical and online classes were conducted in rotation. The same teacher lectured for both groups in Chinese with the same lecture content. Appendix 1 presents a timeline of the course and a list of lectured topics. After the e-Data Protection and File Protection lectures, students were assigned homework. STR technology was used for and by the experimental group only. The teacher informed the experimental group students at the first class that online, one-way lectures will be conducted by using the texts pre-generated by the STR application. In addition, the experimental group was encouraged to use the STR technology to complete homework and identify strengths and limitations of STR.

<table>
<thead>
<tr>
<th>Category</th>
<th>Control group (n=19)</th>
<th>Experimental group (n=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>47.37</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>52.63</td>
</tr>
<tr>
<td>Age (years old)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21–30</td>
<td>1</td>
<td>5.26</td>
</tr>
<tr>
<td>31–40</td>
<td>8</td>
<td>42.10</td>
</tr>
<tr>
<td>41–50</td>
<td>5</td>
<td>26.32</td>
</tr>
<tr>
<td>51 or above</td>
<td>5</td>
<td>26.32</td>
</tr>
</tbody>
</table>

Experimental Design

This study employed a quasi-experimental design following the general recommendations of Creswell (2008). The experiment adopted a nonequivalent control group design to evaluate differences in the control and experimental groups’ learning performance. The study administered a questionnaire survey to investigate students’ perceptions and behavioral intentions toward using STR. The study adopted an independent sample test and effect size methods to test the differences in learning performance of the control and experimental groups in accomplishing homework and post-test objectives. The study conducted one-on-one semi-structured interviews with the experimental group after the experiment to explore the potential effectiveness of applying STR in synchronous cyber classrooms.
Experimental Tools

This study employed the following tools: 1) evaluation of students’ prior knowledge (pretest); 2) assessment of students’ levels of cognitive development (HW1 and HW2); 3) evaluation of students’ learning achievement (post-test); and 4) a questionnaire survey and one-on-one semi-structured interviews regarding students’ perceptions and behavioral intentions toward STR.

The pretest featured 20 multiple choice questions and took place during the first class. The post-test had 10 true or false questions and 10 multiple choice questions, and it took place during the last class. The content of the pretest and the post-test evaluations related to the Information Security course. Both tests were scored on a 100-point scale (with 100 as the highest score), yet the tests were different in content. The study employed Taxonomy for Information Security Education (van Niekerk & Thomson, 2010) adopted from Bloom (1956) to assess homework in order to determine students’ level of cognitive development. The taxonomy (see Appendix 2) includes six levels; each level increases in complexity as the learner moves through the levels. The study adopted a concept as a coding unit and six-point scales for homework assessment. A score of “1” represented the lowest level of cognitive development, and a score of “6” represented the highest level. The final score for homework was the score that corresponded to the highest level of cognitive development found in the homework. For example, if the highest cognitive level in homework was identified as “4” (Analyze), then the homework was scored with a “4.” The assessments were created by a teacher, with more than 10 years of teaching experience in the Information Security domain; thus, the assessments provided superior validity under this condition.

The questionnaire was designed based on the Technology Acceptance Model (Davis, 1986). Four dimensions were covered in the questionnaire: perceived ease (of STR) use (PEU); perceived usefulness (of STR) for learning (PUL); perceived usefulness (of STR) during online one-way lectures (PUOWL); and behavioral intention (BI) to use STR for learning in the future. According to Davis (1986), PEU is the degree to which a student believes that using STR would be free of physical and mental effort. PUL is the degree to which a student believes that using STR for learning would enhance his or her learning performance. PUOWL is the degree to which a student believes that using STR during online one-way lectures would enhance his or her learning performance. BI is hypothesized to be a major determinant of whether or not a student actually uses STR. Responses to the questionnaire items were scored using a five-point Likert scale, anchored by the end-points “strongly disagree” (1) and “strongly agree” (5). Twenty-four valid answer sheets to the questionnaire were obtained out of twenty-five experimental students.

One-on-one semi-structured interviews with subsequent data analysis followed the general recommendations of Creswell (2008). Five students were randomly selected for the interviews. The interviews contained open-ended questions in which students were asked about the following: 1) their experience using the STR application during the experiment; and 2) their opinions about the impact of STR-generated texts for learning. Each interview took approximately 30 minutes; all interviews were audio-recorded with the permission of the interviewee and then fully transcribed for analysis. The text segments that met the criteria for providing the best research information were highlighted and coded. Next, codes were sorted to form categories; codes with similar meanings were aggregated together. Established categories produced a framework to report findings to the research questions.

Statistical Analysis Methods

The study adopted the following methods of statistical analysis:
1. Cohen’s kappa – to evaluate the inter-rater reliability of the assessment (Creswell, 2008; Punch, 2009), i.e., pretest, HW1, HW2, and post-test. The analysis result exceeded 0.72, indicating its high reliability.
2. Cronbach α – to assess the internal consistency of the survey (Creswell, 2008). The value for PEU = 0.89; PUL = 0.94; PUOWL = 0.97; and BI = 0.84, which indicated that the reliability of the items was satisfied.
3. Independent samples test (t-test) – to compare the difference in learning performances for the control and experimental groups (Creswell, 2008) on the pretest, homework, and post-test.
4. The standardized mean difference statistic (referred as d) – Creswell (2008) suggested quantifying the practical strength of the difference between variables through effect size; this approach is important in a quantitative study, especially when using a small sample size to know the significance of a statistical test. He suggested that the effect size of .20 is small, .50 is medium (or moderate), and .80 is large.
Results and Discussion

In this section the results of the study are presented as they relate to each research question and the pedagogical implications.

Research question (1): What are the students’ perceptions and behavioral intentions regarding using STR technology in a synchronous learning environment?

The questionnaire survey data analysis revealed that almost all items in the dimension, “Perceived ease of STR use” were ranked high, as shown in Table 2. This indicates that students generally agreed that STR application was easy to use. However, item number (no.) 2 was ranked as the lowest in this dimension. The interviews with students revealed the reason behind this phenomenon. The students mentioned some particular cases when it was difficult to attain a high recognition accuracy rate of STR. One example is recognizing homophones, the words with the same pronunciation but different meanings. Another example related to the speech of a speaker. When the speaker spoke too slowly, the STR application recognized one spoken word as two. Conversely, when the speaker spoke too quickly, the STR application recognized two spoken words as one (Wald & Bain, 2008; Way et al., 2008).

Table 2. Perceived ease of STR use

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Undecided (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Learning to operate the STR is easy for me.</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>9</td>
<td>4.04</td>
<td>1.08</td>
</tr>
<tr>
<td>2.</td>
<td>I find it easy to get the STR to do what I want it to do.</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>11</td>
<td>2</td>
<td>3.46</td>
<td>0.93</td>
</tr>
<tr>
<td>3.</td>
<td>Interacting with the STR does not require a lot of my mental effort.</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>18</td>
<td>2</td>
<td>3.79</td>
<td>0.83</td>
</tr>
<tr>
<td>4.</td>
<td>My interaction with the STR is clear and understandable.</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>6</td>
<td>4.00</td>
<td>0.78</td>
</tr>
<tr>
<td>5.</td>
<td>It is easy for me to become skillful at using STR.</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>16</td>
<td>3</td>
<td>3.87</td>
<td>0.68</td>
</tr>
<tr>
<td>6.</td>
<td>Overall, I found the STR easy to use.</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>4.08</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Table 3. Perceived usefulness of STR for learning

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Undecided (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Using STR improves the quality of my learning.</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>6</td>
<td>4.08</td>
<td>0.65</td>
</tr>
<tr>
<td>8.</td>
<td>STR helps me to accomplish learning tasks more quickly.</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>13</td>
<td>6</td>
<td>4.04</td>
<td>0.69</td>
</tr>
<tr>
<td>9.</td>
<td>Using STR increases my productivity.</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>3.62</td>
<td>0.77</td>
</tr>
<tr>
<td>10.</td>
<td>Using STR enhances my effectiveness for learning.</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>3</td>
<td>3.62</td>
<td>0.87</td>
</tr>
<tr>
<td>11.</td>
<td>Using STR improves my learning achievement.</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>3.58</td>
<td>0.83</td>
</tr>
<tr>
<td>12.</td>
<td>Overall, I find STR useful in my learning.</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>13</td>
<td>7</td>
<td>4.12</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Table 3 shows that almost all items in the dimension, “Perceived usefulness of STR for learning” were ranked high. This demonstrates that most students agreed about how useful STR technology was for learning (Ryba et al., 2006; Wald, 2010; Wald & Bain, 2008). Only item no. 11 was ranked as the lowest in this dimension. The interviews with the students helped infer such a finding. Some students, for various reasons, did not attend online classes; they preferred to make up missed classes by studying the learning materials from textbooks or via the Internet instead of from recorded archives of online classes. These students had no idea STR-generated texts could be used during and after online classes, nor did they realize the benefits of using STR-generated texts for learning.

All items regarding the dimensions, “Perceived usefulness of STR in one-way lectures,” shown in Table 4, and “Behavioral intention to use STR,” shown in Table 5, were ranked with high scores. Most students agreed that STR-generated text during one-way lectures was useful for learning, and most students were highly motivated to use STR continuously after this study.

Table 4. Perceived usefulness of STR in one-way lectures

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Undecided (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Use of STR by the teacher during one-way lectures improves the quality of my learning</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>14</td>
<td>4.37</td>
<td>0.97</td>
</tr>
<tr>
<td>14</td>
<td>Use of STR by the teacher during one-way lectures helps me to accomplish learning tasks more quickly</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>11</td>
<td>4.25</td>
<td>0.94</td>
</tr>
<tr>
<td>15</td>
<td>Use of STR by the teacher during one-way lectures increases my productivity</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>4.00</td>
<td>0.98</td>
</tr>
<tr>
<td>16</td>
<td>Use of STR by the teacher during one-way lectures enhances my effectiveness in learning</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>11</td>
<td>4.20</td>
<td>0.98</td>
</tr>
<tr>
<td>17</td>
<td>Use of STR by the teacher during one-way lectures improves my learning achievement</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>3.92</td>
<td>1.06</td>
</tr>
<tr>
<td>18</td>
<td>Overall, I found using STR by the teacher during one-way lectures is useful in my learning</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>13</td>
<td>4.42</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 5. Behavioral intentions to use STR

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Strongly disagree (1)</th>
<th>Disagree (2)</th>
<th>Undecided (3)</th>
<th>Agree (4)</th>
<th>Strongly agree (5)</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>I intend to continue using STR in the future</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>11</td>
<td>4.29</td>
<td>0.75</td>
</tr>
<tr>
<td>20</td>
<td>I plan to use STR often.</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>4.04</td>
<td>0.81</td>
</tr>
<tr>
<td>21</td>
<td>I will strongly recommend the use of STR to others.</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>4.25</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Research question (2): Do the students using STR-generated texts perform better in accomplishing homework and post-test objectives than the students who do not use the STR technology?

Table 6 shows the means and standard deviations of students’ scores on the pretest, homework, and post-test and the results of the t-test and effect size. According to the results, all mean values of the experimental group were higher than the mean values of the control group. The results demonstrate that many assessment items had large standard deviations values. The results of the t-test showed no significant difference in the performance of the two groups on the pretest ($t=-2.294$, $p=0.034$). However, the results demonstrated that the students of the experimental group significantly outperformed the students of the control group on the post-test ($t=-2.239$, $p=0.034$). An effect size of 0.03 was obtained for the pretest, 0.50 for HW1, 0.55 for HW2, and 0.70 for the post-test. The results revealed that the average student in the experimental group would perform higher than a student in the control group as follows: over 0.5 standard deviations in HW1 achievement; over 0.55 standard deviations higher in HW2 achievement; and over 0.7 standard deviations higher in post-test achievement.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Control</th>
<th>Experimental</th>
<th>$t$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>73.42</td>
<td>14.15</td>
<td>74.80</td>
<td>16.30</td>
</tr>
<tr>
<td>HW1</td>
<td>4.21</td>
<td>1.47</td>
<td>4.84</td>
<td>0.99</td>
</tr>
<tr>
<td>HW2</td>
<td>3.79</td>
<td>1.18</td>
<td>4.40</td>
<td>1.04</td>
</tr>
<tr>
<td>Post-test</td>
<td>70.79</td>
<td>21.75</td>
<td>83.20</td>
<td>12.07</td>
</tr>
</tbody>
</table>

* $p<0.05$

Large standard deviation values obtained in this study revealed a diversity in learning performance. One possible reason for this finding could be that the participants were Open University students and they have a large variety as far as careers were concerned, interests, and learning abilities. This phenomenon should be further investigated in the future. The results of the t-test showed that the two groups of students had similar levels of prior knowledge at the first week of the experiment, and similar levels of knowledge about the subject at the time homework was performed (week 7 and week 9). This suggests that the majority of experimental students could not fully use STR-generated text until homework was assigned. And even once they started using STR-generated texts to do homework, it was their first experience of using texts to increase their learning capacities. However, the effect size results demonstrated that some students of the experimental group, although not all, performed moderately better than the students of the control group on homework. This implies that some experimental students did use STR-generated texts efficiently, and hence enhanced their understanding of the lectures’ content, benefited from the texts while doing homework after the lecture (Ryba et al., 2006; Wald, 2010; Wald & Bain, 2008), and performed better on HW1 and HW2. The results of the t-test and the effect size revealed that the experimental group performed significantly better than the control group on the post-test, at the last week of the experiment (week 12). The results suggest that most of the experimental students eventually started using STR-generated texts after they have sensed that texts could increase their ability to learn. The students got more experienced using STR-generated texts during the lectures and doing homework. The results also indicate that the students intensively studied STR-generated texts before the post-test to obtain higher scores. Subsequently, the experimental group performed significantly better compared to the control group. Based on these findings, we could conclude that the use of an STR application has more beneficial effects on students’ learning if the period of using STR-generated texts is extended, and if students can identify its strength.

Research question (3): How were the STR-generated texts beneficial to students’ learning and what reasons were deduced from interviews?

During the one-on-one semi-structured interviews, students mentioned that they could benefit from STR-generated texts during an online, one-way lecture, after it, and while doing homework. Regarding the use of STR technology during online one-way lectures, the students pointed out technical problems that occasionally occurred in the synchronous cyber classroom. The interruption of the teacher’s voice and unclear audio were issues caused by an unstable Internet connection or an unsettled microphone, which could hinder the students’ understanding of oral lectures (Chen et al., 2005; Hastie et al., 2010; Wang et al., 2010). The students emphasized the usefulness of STR-generated texts in complementing these problems. The following content was pulled from two interviews.
I usually engaged myself in listening to the lecture, but when I encounter some problems like an unclear lecture, an unstable Internet connection, or interference from other events, I miss some information. I can read the text generated by STR instead.

Due to some problems on the Internet, I experienced unclear speech by the teacher during an online, one-way lecture when a big number of students attended it. Likely, STR could help overcome such problem.

In addition, the students preferred to listen to online lectures and read the STR-generated text simultaneously during online, one-way lectures. The reasons for such behaviors were that the students wanted to make up missed information, clarify lecture content, and enhance their understanding of the teacher’s lecture in a synchronous learning environment. The students also mentioned that STR-generated texts are useful in simultaneous note-taking; the students took notes and reviewed them later to get a deeper understanding of the lecture content (SRS, 2011; Wald & Bain, 2008). The following content is derived from two interviews.

When I paid attention to the lecture, I read the STR text at the same time. I would like to double-check and make sure that there was no misunderstanding of the content. ...It is like you listen to the class and someone takes notes for you at the same time. Thus, it’s easy to understand the lecture.

Text generated by STR during online lectures helps me understand the course content. It is common to miss some information during the class due to some reasons, but with the assistance of STR text, I feel more relaxed in class and I don’t have to worry about missing any important information.

Regarding the use of STR-generated texts after online, one-way lectures, the students studied lecture transcripts to make up the classes they missed, while some students studied transcripts to get a better understanding of lecture content (Wald, 2010; Way et al., 2008). The students favored the STR-generated text over other media (e.g., video, audio, and PowerPoint® slides); it took less time to study the texts. The students also mentioned that STR-generated texts were useful for post-lecture note-taking and for doing homework. They could copy and paste excerpts from STR-generated texts into their notes or homework and then revise the excerpts to make final notes or homework. The following content is the result of two interviews.

I used text generated by STR after online lectures. Mainly I did it for homework, but I also used to taking notes to grasp some key concepts from STR text after class, instead of seeing the video lecture.

I often work late, so I am usually late for the online lectures. Then I miss some parts of the online lectures, but STR text helps me to catch up on that information. It’s easy for me to review and complete homework using text generated by STR.

Pedagogical Implications

Based on obtained findings, this study makes the following implications and recommendations for educators who plan to teach in online synchronous learning environments using STR technology. First, we recommend that students use STR-generated text to enhance their understanding of a teacher’s lecture content in an online synchronous cyber classroom as well as take advantage of STR-generated text both during and after lectures. Second, the “training” of STR technology by educators is necessary prior to obtaining a reasonable technological accuracy rate (<85 percent) (Wald & Bain, 2008; Way et al., 2008). This study recommends training with STR technology using training scripts with content related to the Information Security course so that the STR technology can “learn” domain-specific terminology during the training period and it can recognize words with fewer errors after training (during actual online lectures). Third, this study recommends several rehearsals of online lectures using STR technology before actual lectures take place. The rehearsals can help educators gain experience working with STR technology as well as to become adapted to a speech pattern that best “suits” the STR technology (with moderate speed and volume, less spontaneity, fluency, etc.). The texts generated by STR during the rehearsals can also be used by the educators for one-way lectures, namely, to replace the original texts inaccurately generated by STR from online lecturing. In this case, a teaching assistant should be in charge if the task of displaying the text during the one-way lectures; it will reduce the teacher’s burden and be more focused on lecturing for a smoother lecture, thus avoiding delays in
transcription. Forth, this study recommends that educators prepare notes with the main points of the lecture. These notes can be helpful during online lectures in adjusting the educator’s speech to a less spontaneous mode; therefore, text can be generated from STR technology with a higher accuracy rate. Furthermore, this study recommends that educators design STR technology-based teaching and learning activities in a way that encourages students to find the strengths and limitations of STR technology. The previously mentioned recommendations were used in this study by the teacher, and as a result, STR-generated texts achieved a high accuracy rate and the texts became useful and meaningful for students’ learning both during and after lectures.

Conclusion

The study was conducted with appropriately designed learning activities by using STR technology in an online synchronous lecturing environment as well as in a situation wherein an individual student completed homework. The aim of this study was to investigate the students’ perceptions and behavioral intentions toward using STR and the effectiveness of applying STR in synchronous cyber classrooms. The study revealed that the students in the experimental group perceived the STR technology to be easy to use and useful for one-way lectures and individual learning. Most students mentioned that they were highly motivated to use STR to learn in the future. The results of effect size revealed moderate improvement in performance of the experimental group over the control group on homework accomplishments. The results of the *t*-test and effect size showed that the students in the experimental group performed significantly better on the post-test than the students in the control group. These results imply that the students in the experimental group used STR-generated texts during one-way lectures and for individual learning; thus, the students benefited from STR-generated texts to enhance their overall learning experiences. The interviews with the students revealed that STR-generated texts were used by the students as study tools to clarify lecture content and to take notes during one-way lectures. After the online lectures, the students benefited from STR-generated text by studying it to recall and to better understand previous one-way lectures, to take notes, and to reflect on what they learned during the lecture by doing homework. Previous studies show similar findings regarding using the STR-text during and after one-way lectures (Ryba et al., 2006; SRS, 2011; Wald & Bain, 2008; Way et al., 2008). Based on the obtained findings, this study suggests that educators use STR mechanisms to support teaching and learning activities in synchronous learning environments and to encourage students to use STR-generated texts both during and after lectures to facilitate learning.

Several limitations need to be acknowledged and addressed regarding this study. The first limitation is the relatively small sample size, which limits the broad generalization of the results. The second limitation is the consequent adoption of STR by the teacher and students. It requires the teachers to take time to train the system and to edit STR-generated text before or after classes; thus, a teacher's voluntarily adoption may be very problematic. The third limitation pertains to the accuracy rate of STR technology and the delay of text generation during online, one-way lecturing, which restricts full usage of STR mechanisms. The last limitation relates to large standard deviation values, which suggests that students’ scores were spread out over a large range of values. Future studies should be conducted to investigate related issues. One promising theory that needs to be investigated in the future is the pedagogical effectiveness of using STR technology to support other learning activities, like group discussions, in synchronous learning environments. For example, we believe that using STR technology for group discussions in synchronous learning environments is useful to improve students’ learning, particularly when students use STR-generated text for deeper reflection and summarization. Furthermore, students’ perceptions and acceptance of using STR technology for group discussions needs to be investigated. Meanwhile, advanced applications using STR mechanisms to conduct different language communications or lectures in traditional classroom environments are also worth further investigation.

Acknowledgements

This work was supported in part by the National Science Council of Taiwan, R.O.C. under grant NSC 99-2631-S-006-001, NSC-98-2631-S-008-001, NSC 98-2511-S-008-008-MY3, and NSC 98-2511-S-008-005-MY3 and the Research Center for Science & Technology for Learning of the University System of Taiwan.
References


APPENDIX 1. Timeline of the course and topics of the lectures

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Topics of the lectures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Pretest</td>
</tr>
<tr>
<td></td>
<td>Introduction to information security (F2F*)</td>
</tr>
<tr>
<td>Week 2</td>
<td>Risk management (OL**)</td>
</tr>
<tr>
<td>Week 3</td>
<td>Computer crime, ethics and law (F2F)</td>
</tr>
<tr>
<td>Week 4</td>
<td>Information security standards (OL)</td>
</tr>
<tr>
<td>Week 5</td>
<td>Break</td>
</tr>
<tr>
<td>Week 6</td>
<td>Network and wireless communication security (F2F)</td>
</tr>
<tr>
<td>Week 7</td>
<td>e-Data Protection (OL)</td>
</tr>
<tr>
<td></td>
<td>HW1 assignment</td>
</tr>
<tr>
<td>Week 8</td>
<td>Access control (F2F)</td>
</tr>
<tr>
<td></td>
<td>HW1 submission deadline</td>
</tr>
<tr>
<td>Week 9</td>
<td>File Protection (OL)</td>
</tr>
<tr>
<td></td>
<td>HW2 assignment</td>
</tr>
<tr>
<td>Week 10</td>
<td>Email and Web security (F2F)</td>
</tr>
<tr>
<td></td>
<td>HW2 submission deadline</td>
</tr>
<tr>
<td>Week 11</td>
<td>Physical and Operation security (OL)</td>
</tr>
<tr>
<td>Week 12</td>
<td>Computer virus and course summary (F2F)</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
</tr>
</tbody>
</table>

*F2F-face to face lecture; **OL-online lecture
APPENDIX 2. Taxonomy for Information Security Education

<table>
<thead>
<tr>
<th>Level</th>
<th>Category</th>
<th>Assessment criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Remember</td>
<td>What is the definition of information leakage? What is computer virus?</td>
</tr>
<tr>
<td>2</td>
<td>Understand</td>
<td>Summarize the given security policy in your own words. Explain how the password cracking works.</td>
</tr>
<tr>
<td>3</td>
<td>Apply</td>
<td>Think of three things that could go wrong should your password be compromised. Follow the guideline to install the PGP tool and create your own asymmetric key pair.</td>
</tr>
<tr>
<td>4</td>
<td>Analyze</td>
<td>Which of the following network architectures are most dangerous? Compare and contrast the security functionality of firewall to intrusion detection system.</td>
</tr>
<tr>
<td>5</td>
<td>Evaluate</td>
<td>Which of the following encrypt algorithms would be more robust? Why? Is it fair for a company to insist that employees never use their work email for personal matters?</td>
</tr>
<tr>
<td>6</td>
<td>Create</td>
<td>Pretend you are an information security officer for a large firm. Write a report about a recent security incident. Rewrite a given incident report as a news story.</td>
</tr>
</tbody>
</table>
Teachers’ Belief and Use of Interactive Whiteboards for Teaching and Learning

Yalın Kilç Türel and Tristan E. Johnson

Department of Computer Education and Instructional Technology, Firat University, Elazig, Turkey // Learning System Institute, Florida State University, Tallahassee, Florida, USA // yturel@gmail.com // tejohnson@fsu.edu

ABSTRACT

Interactive whiteboards (IWB) are regarded as one of the most revolutionary instructional technologies for various educational levels. While the impacts of IWBs in classroom settings have been examined recently in a number of studies, this study not only looks at the perception but also examines the actual usage and behaviors associated with promising IWB features in practical settings. The main goal of this paper is to evaluate both teachers’ perceptions and their use of IWBs. A questionnaire was developed based on an extensive literature review as well as related instructional theories and models. The questionnaire consisted of questions about demographics, usage, and teachers’ perceptions related to IWBs. For this study, 174 teacher-participants, who have actively used IWBs for instruction, were selected from various educational levels (from grade 6 to 12). The results show that teachers believe that IWBs can be used for different subject domains. Also, teachers believe that IWBs can be used to facilitate learning and instruction under the following conditions, 1) collaboration with colleagues, 2) training about effective instructional strategies using IWB, and 3) more frequent teacher use of IWBs to improve IWB competency.

Keywords
Interactive whiteboard, Interactive whiteboard teacher scale, Interactive learning environments, Secondary education, Technology acceptance model, Technology integration, Technology adoption

Introduction

Educational institutions have tried to provide students better learning environments by equipping them with the latest technology. This effort has encouraged instructors to use various assistive technologies such as computers and the Internet in their classrooms especially over the last decade; this process is called integration of information and communication technologies (ICT) (Hsu, 2010). As a part of the ICT integration process, the interactive whiteboard (IWB) has been one technology most invested in especially by European countries such as England, Spain, and Turkey (Holmes, 2009; Türel, 2010). As of 2010, England has the highest IWB penetration rate (73%) in the world and many countries including Denmark (50%) and the USA (35%) have substantially increased IWB rates in classrooms; however, the average rate for Asia is still lower than 2% according to a recent research report (McIntyre-Brown, 2011).

IWB’s promising benefits to learning and instruction have led to its increased popularity and attractiveness as expressed by a number of researchers (Bell, 2002; Levy, 2002; BECTA, 2003; Brown, 2003; Beauchamp & Parkinson, 2005; Smith, Higgins, Wall, & Miller, 2005; Slay, Siebörger, & Hodgkinson-Williams, 2008). It is asserted that IWBs can enhance the functionality of existing ICT such as computers and projectors by adding interactivity to these media that make it distinct from traditional PowerPoint presentations (Hall & Higgins, 2005; Smith et al., 2005; Torff & Tirotta, 2010). Considering the possible advantages of IWBs, teachers can enrich their instructions with various instructional strategies and techniques and, therefore, increase students’ attention, motivation, participation, and collaboration by means of an IWB (Levy, 2002; Beauchamp & Parkinson, 2005; Hall & Higgins, 2005; Glover, Miller, Averis, & Door, 2007). Although those researchers strongly emphasize the positive effects of this technology when appropriately integrated into classrooms, the true success of IWBs depends on how they have been used by teachers in a learning context. Teachers report that they have used IWBs through appropriate instructional strategies and resources, and above all, learned to incorporate the IWB with their course content. Hence, researchers, who have attempted to evaluate IWB use, have relied on perceptions of teachers as the main data source (Slay, Siebörger, & Hodgkinson-Williams, 2008), to determine the effectiveness of this technology in school settings.

Instructional use of interactive whiteboards

The interactive whiteboard (IWB) has been used in many contexts in various ways. Effective use of IWBs in classrooms promises numerous advantages in terms of learning and instruction (Türel, 2010). However, to better
understand how we can effectively use IWBs in classrooms, several questions should be considered: What instructional strategies can teachers use with IWBs and what kinds of benefits do IWBs have on teaching and learning.

Successful instruction may be a result of various IWB features along with sound instructional strategies (Brown, 2003; Glover et al., 2007). Teachers can put a variety of strategies and techniques into practice using IWBs by considering the characteristics of the learning context including students’ needs and interests, and technical facilities (Türel, 2010). Several IWB instructional strategies that have a positive effect on student learning include:

- Highlighting, coloring, or annotating important content (Türel & Demirli, 2010)
- Flipping back and forth to review previous content providing reviewing techniques better understanding (Levy, 2002; Smith et al., 2005)
- Using pictures for discussion, collaborative writing, shared reading, peer-teaching, and collaborative problem solving (BECTA, 2006)
- Hiding and reveal, drag and drop, and matching items activities (Türel, 2010)
- Observing different media—essential for visual learners (Bell, 2002)
- Touching and feeling the material—good for tactile learners (Bell, 2002)
- Accommodating lower ability and special needs—zoom feature for visually impaired students (Smith, 2008)
- Presenting ideas and reflections about the course content
- Finding hidden part of a picture with spotlight or screen-shade (Beauchamp & Parkinson, 2005)
- Capturing screenshots from web pages synchronously and manipulating them
- Correcting mistakes in the materials (Beauchamp & Parkinson, 2005)
- Playing games (Smith et al., 2005)

Benefits of IWB technology include:

- Enhanced social interaction (Türel & Demirli, 2010)
- Reformed learning environments—teachers may facilitate student’s involvement, interaction, and collaboration (Smith et al., 2005)
- Draw the learners’ attention (Türel, 2010)
- Facilitated learning and remembering using visual media (Türel, 2010).
- Enlarged computer touch screen
- Interactions can be recorded and saved—Acrobat (PDF) document, PowerPoint slides, or record whole lecture as a movie file
- Using with voting systems, document cameras, and electronic microscopes (Bell, 2002)

To this end, as well as having IWB technical competencies and skills, teachers should also be aware of such pedagogical implications in order to provide effective instruction to their students using IWBs (Türel, 2010). Although research suggest that an ideal use of IWBs may have a positive impact on learning and instruction, it is important to investigate how teachers in classroom settings are using IWBs.

In order to better understand teachers’ IWB use, the examination of different factors is needed such as time, instructional strategies, and techniques. Depending on the frequency and duration of IWB use, teachers gradually develop their skills and abilities (Hodge & Anderson, 2007). However, overuse of IWBs as a presentation tool in a teacher-led instructional setting may deteriorate students’ motivation, attention, and consequently, the efficiency of instruction (Hall & Higgins, 2005).

Results of studies in various contexts such as different countries, across educational levels, and subject domains, demonstrate teachers’ positive perceptions about IWBs (Slay, Siebörger, & Hodgkinson-Williams, 2008). However, in some studies, teachers reported several IWB issues that may dramatically decrease the effectiveness of IWBs in their courses (Somyürek, Atasoy, & Özdemir, 2009). As a crucial issue, many studies (Levy, 2002; BECTA, 2003; Smith et al., 2005) address teachers’ insufficient IWB knowledge and experiences. One solution is to provide in-service training focusing on effective IWB strategies. Several studies (Glover & Miller, 2001; Smith et al., 2005) stress that teachers often get limited IWB training from suppliers that only covers basic IWB skills. Likewise,
teachers may improve their IWB skills by themselves or by collaborating with colleagues (Shenton & Pagett, 2007). Providing an appropriate training program depends on the assessment of needs, problems, expectations, and skill levels of IWB users. To this end, Beauchamp (2004) developed a framework that classifies IWB users based on behaviors and features most often used. Beginners use IWBs as a traditional blackboard, while advanced users use IWBs to construct meaning using interactive and fluid lesson strategies. Technical competencies are examined looking at navigating an operation system, importing media from different sources, properly using hyperlinks between programs, and regularly saving lessons (Beauchamp, 2004). Based on this framework, it was found that higher-level experienced teachers use more IWB features.

Similar to this correlation, increasing use of technology is strongly correlated to teachers’ acceptance and positive attitudes about the technology. Thus, the duration or frequency of teachers’ IWB use is regarded as another correlating factor that may impact teachers’ perceptions about IWBs. However, it is impossible to measure the exact IWB exposure time during instruction due to the variety of IWB techniques that teachers use. Some strategies only use IWBs while others incorporate a combination of techniques such as IWB with group discussions for example. Even within an IWB focused strategy, teachers may use lecture and also demonstration along with drill and practice. As such, exposure time had a relative effect because teachers use IWBs for different purposes (Levy, 2002). In addition, the amount of elapsed time of IWB use based on hours in a week (e.g., Erduran & Tataroğlu, 2009) may be misleading because the total hours some teachers work in a week may vary. Thus, Moss et al. (2007) examined the frequencies of teachers’ IWB use under the following categories: never, hardly ever, some, most, and every lesson. They found no significant differences between Mathematics, Science, and English teachers in terms of the frequency of IWB use. However, in reducing the frequencies to low and high IWB use, Moss et al. (2007) found that math teachers in the high IWB use group revealed stronger positive attitudes towards IWBs than the low IWB use group.

Significance of the study

In many studies (e.g., Bell, 1998; Beeland, 2002; Cogill, 2002; Levy, 2002; Beauchamp, 2004; Wall, Higgins, & Smith, 2005; Moss et al., 2007), teachers’ preferences, needs and perceptions about IWBs have been examined to better understand how these factors impact IWB use. The majority of those studies investigated teachers’ perceptions by means of questionnaires focusing on particular variables such as attitudes (Beeland, 2002), motivation (Wall, Higgins, & Smith, 2005; Torff & Tirotta, 2010), satisfaction (Bell, 1998), interaction (Levy, 2002; Glover et al., 2007), acceptances (Saltan, Arslan, & Gök, 2010), and technical issues of IWB use (Wall, Higgins, & Smith, 2005; Somyürek, Atasoy, & Özdemir, 2009).

In addition to perceptions, current practices need to be measured and analyzed to best understand teachers’ technology use. In the literature, it seems there are several paucities and limitations regarding overall IWB evaluation since they are primarily based on teachers’ perceptions. Some studies only focused on a specific discipline area (e.g., Glover et al., 2007) while others have a limited sample size (e.g., Bell, 1998; Beeland, 2002; Tozcu, 2008). Additionally, teacher-participants who have not used or have just started to use IWBs in their classes may be an inappropriate data source for an evaluation of perceptions because of their insufficient knowledge, experience, and attitudes about the use of IWBs. Another critical issue for IWB research is the use of appropriate surveys and questionnaires that were developed based on existing research as well as sound instructional theories and strategies associated with the use of IWBs. Addressing the issues regarding the evaluation of IWB use, we conducted this study by means of an original instrument on a particular group of IWB-experienced teachers who were teaching at different educational levels and various disciplines in Turkey.

Bearing in mind the increase of IWB technology investments, there is a strong need for the evaluation and thus, improvement of actual IWB use in schools (Slay, Siebörger, & Hodgkinson-Williams, 2008). Considering the importance of teachers’ perceptions, attitudes and beliefs about IWB use, this study focuses on the multiple component investigation of IWBs based on teachers’ perceptions on their current IWB use including frequencies of usage, preferred IWB features, status quo of IWB skills and training as well as perceptual benefits of IWB in classroom teaching and learning. Based on the purpose of this study, the research questions are as follows:

1. What are the main sources of IWB training for teachers?
2. What IWB training topics do teachers need?
3. How much is each IWB feature being used?
4. What are the teachers’ perceptions about their IWB use?
5. Is there a relationship between teachers’ IWB use frequencies and 1) self-reported competencies, 2) discipline areas, and 3) teachers perceptions?

**Method**

For this study, a quantitative descriptive research method was employed to investigate the perceptions of teachers regarding the current state of IWB use in schools. Descriptive research methods are one of the most preferred and effective methods to depict and interpret the understanding of participants’ beliefs about a certain issue or phenomenon (Gall, Gall, & Borg, 2003). Given that we needed to collect data from many people, it was out of scope for this study to visit each class and observe the use of IWB for each teacher. The potential for self-report bias was minimized since the data was anonymous and was not shared with anyone who directly knew the participants (Chan, 2009). Data was collected from teachers via a questionnaire developed specifically for this study.

**Participants**

The sample for this study consisted of 174 Turkish teachers, ranging from grades six to twelve, who have actively used IWBs for at least six months in their schools. We selected the volunteer participants based on the purposeful and convenience sampling method. Since the focus of this study is to evaluate how teachers use IWBs rather than why teachers do not use IWBs, it was essential to select participants among ones who have had sufficient knowledge of and experience with IWBs and also were familiar with the issues of IWBs in practice. All participants were college educated and 59.8% of them were male. The majority of participants were less than 36 years old (90.8%) and the majority of all the teachers (88.5%) had been teaching for less than ten years. Teachers’ responses were examined in terms of their fields of teaching by categorizing them into six areas: Computer Science, Foreign Language (English), Mathematics, Science, Social Sciences, and Turkish Language and Literature.

**Instrument Development**

In order to create a questionnaire consistent with the study’s purpose, we examined current studies looking at instructional theories and strategies, current practices, problems and perceptions of IWB users (Bell, 1998; Beeland, 2002; Cogill, 2002; Beauchamp, 2004; Wall, Higgins, & Smith, 2005; Moss et al., 2007). The initial draft of the questionnaire was distributed for feedback from 10 teachers who were active IWB users across various subject areas, two instructional designers, two language teachers, and two educational science teachers. Revisions were made based on expert opinions. This step was vital to achieve a comprehensible and relevant questionnaire in terms of face and content validity (Black & Champion, 1976).

As well as demographics and multiple-choices items, the final questionnaire (α = .93) included 26 Likert scale items from strongly disagree to strongly agree. We also classified the Likert scale items along with the existing literature into themes to provide a better understanding of main dimensions of IWB use. Those themes are labeled as instructional effects of IWBs (α = .86), motivational effects of IWBs (α = .89), and the usability of IWBs (α = .60). The first theme includes items related to effects of IWBs on teaching and learning while the second theme has items addressing the motivational issues of IWBs. The last theme includes three items concerning the usability of IWBs. Original language of the paper-based questionnaire delivered for this study is Turkish.

**Analysis**

Considering the goals for the study, a descriptive analysis was performed to understand the current state of teachers’ IWB use as well as teachers’ general perceptions about using IWBs. To provide a clear picture, percentages of teachers’ agreement levels are presented in two groups: agreeing (agree and strongly agree options), and disagreeing (disagree and strongly disagree options). For internal consistency and reliability, Cronbach’s Alpha coefficients were calculated and interpreted for each theme based on the rules (.9 = high level, .8 = moderate, .7 = low level, .6 = acceptable level, and <.6 = unacceptable level) (Murphy & Davidshofer, 1991).
Chi-square tests of independence were performed to analyze the relationships of key categorical variables such as the frequency and duration of IWB use with teachers’ fields, IWB competencies, and their perceptions. Cramer’s V-values were examined for the effect size of associations in accordance with the intervals presented by Kotrlik and Williams (2003).

**Results**

Results for the study are presented in three parts: 1) statistical results of teachers’ IWB use, skills, and training, 2) descriptive results of teachers’ responses to the questionnaire items, 3) results focusing on individual differences between IWB usages and perceptions.

**Teachers’ IWB use and training**

In the first section of the IWB questionnaire, teachers were asked several questions about their use of IWBs in their courses (see Table 1).

<table>
<thead>
<tr>
<th>Table 1. Teachers’ IWB usage statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>How long have you used an IWB (number of year)?</td>
</tr>
<tr>
<td>Less than 1 year</td>
</tr>
<tr>
<td>1-3 year</td>
</tr>
<tr>
<td>More than 3 year</td>
</tr>
<tr>
<td>How many hours do you use IWBs in a week?</td>
</tr>
<tr>
<td>&lt; 3 hour</td>
</tr>
<tr>
<td>4-5</td>
</tr>
<tr>
<td>6-7</td>
</tr>
<tr>
<td>&gt; 7 hours</td>
</tr>
<tr>
<td>Frequency of IWB use</td>
</tr>
<tr>
<td>Sometimes</td>
</tr>
<tr>
<td>Frequently</td>
</tr>
<tr>
<td>Always</td>
</tr>
<tr>
<td>How competent you are as an IWB user?</td>
</tr>
<tr>
<td>1 (incompetent)</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5 (professional)</td>
</tr>
</tbody>
</table>

All participants stated that they had a portable IWB (Mimio or E-beam) in their classrooms. A small number of teachers (5.2%) reported using IWBs less than one hour per week while the majority of teachers (61.5%) reported using IWBs more than seven hours per week. When asked to select one of three levels for their IWB use (sometimes, frequently, or always), teachers mostly selected either ‘frequently’ (42.5%) or ‘always’ (36.8%) while ‘sometimes’ had the lowest rate (20.7%). Using an IWB skills level framework (Beauchamp, 2004), teachers graded themselves as IWB users from incompetent (1) to professional (5). The majority of teachers described themselves as either average (39.1%) or just above average (45.4%) levels while a few teachers acknowledged themselves as either professional (4.6%) or incompetent user (4.6%).

The relationships of key categorical variables were analyzed such as the frequency and duration of IWB use with teachers’ fields, IWB competencies, and teachers’ perceptions. Teachers who did not use IWBs were not included in this study. Chi-square tests showed significant results in the association of IWB use frequencies 1) with teachers’ self-reported IWB competencies \( \chi^2 (8, N = 174) = 47.43, p < 0.05, \text{Cramer's } V^2 = .369 \) and 2) with their discipline areas \( \chi^2 (10, N = 174) = 34.86, p < 0.05, \text{Cramer's } V^2 = .317 \). For the distribution of weekly hours indicating the duration of IWB usage, chi-square tests were performed for two levels (up to seven hours, more than seven hours). Similar to the results of usage frequencies, significant differences were found between 1) the duration of weekly IWB use and teachers’ IWB competencies \( \chi^2 (4, N = 174) = 16.56, p < 0.05, \text{Cramer's } V^2 = .369 \) as well as 2) teachers’ discipline areas \( \chi^2 (5, N = 174) = 24.92, p < 0.05, \text{Cramer's } V^2 = .378 \). These results show that there are positive correlations between the increase on the IWB usage and IWB competencies. Also, there is a positive...
correlation between teachers’ IWB usages and their fields. All significant relationships demonstrate a moderate association based on the interpretation of Cramer’s $V$-values between 0.2 and 0.4 (Kotlrik & Williams, 2003).

A multiple-selection question was used to determine how teachers use the most prominent IWB features. Teachers could select any feature for this question with two options: seldom and frequently (see Figure 1).

![Figure 1. Usage frequencies of IWB features](image)

The results show how often the features are used as well as total frequencies. Participants most commonly use IWB’s mouse function while the fewest number of teachers (N=33) used hyperlinks.

Teachers were asked about how they obtained their IWB skills and knowledge. Four options were posed to teachers and they were instructed to select up to two options (Figure 2). Restraining the options was designed to extract the most important training sources for teachers.

![Figure 2. Main sources of IWB training for teachers](image)
Results show that while 164 respondents marked at least one option to this question, 90 (55%) respondents reported that they received training from either their own institutions or the vendor of the IWB while 20 (12%) responded that they were trained by both sources. In addition, 76 respondents reported that they learned to use IWBs from their colleagues while 26 teachers reported that they learned using IWBs by themselves.

Given three main IWB topic categories, teachers were asked about what IWB training topics they had taken, what topics were needed, or what topic they did not need to take. The topics included (1) technical IWB information and skills, (2) effective teaching methods and techniques for an effective IWB use, and (3) finding and designing instructional materials compatible with IWBs (Figure 3).

The results show that there are a large number of teachers that say that they do not need to take training related to the topics.

**Teachers’ Responses to the Questionnaire Items**

The results of teachers’ responses to the 26 Likert-scale items in the questionnaire were examined according to three main themes: (a) instructional effects of IWBs, (b) motivational effects of IWBs, (c) the usability of IWBs.

**Instructional Effects of IWBs**

Teachers responded to the questions related to the instructional effects of the IWB use on teaching and learning (Table 2). Cronbach’s alpha coefficient for this part of the questionnaire is 0.86.

<table>
<thead>
<tr>
<th>Statements</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>% of teachers disagreeing/agreeing with each statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>In terms of teaching</td>
<td></td>
<td></td>
<td></td>
<td>Disagree</td>
</tr>
<tr>
<td>Q1. IWB helps me to manage instructional time effectively</td>
<td>172</td>
<td>4.27</td>
<td>.943</td>
<td>5.8</td>
</tr>
<tr>
<td>Q2. I think the lessons become more effective with IWB</td>
<td>168</td>
<td>3.90</td>
<td>1.010</td>
<td>10.1</td>
</tr>
<tr>
<td>Q3. IWB facilitates the classroom management for me</td>
<td>158</td>
<td>3.77</td>
<td>.944</td>
<td>9.5</td>
</tr>
<tr>
<td>Q4. IWB helps my lessons be more interactive</td>
<td>157</td>
<td>3.82</td>
<td>.859</td>
<td>7.6</td>
</tr>
<tr>
<td>Q5. IWB facilitates discussions on the content in class</td>
<td>158</td>
<td>3.30</td>
<td>.954</td>
<td>17.7</td>
</tr>
<tr>
<td>Q6. There is no time for my students to get around to using an IWB*</td>
<td>162</td>
<td>2.99</td>
<td>1.098</td>
<td>37.0</td>
</tr>
</tbody>
</table>
Q7. IWB provides advantages to me to make course content more visual
Q8. The way I give instruction has been changed since I began to use an IWB
Q9. IWB helps me to use the computer and projector more effectively than before

In terms of learning
Q10. I believe using an IWB helps my students’ learning
Q11. Using an IWB makes it easier for my students to remember what they learned in class
Q12. My students learn faster when I teach with an IWB
Q13. IWB helps my students to learn in groups
Q14 Using an IWB helps students to learn concepts easier

In terms of learning
Q10. I believe using an IWB helps my students’ learning
Q11. Using an IWB makes it easier for my students to remember what they learned in class
Q12. My students learn faster when I teach with an IWB
Q13. IWB helps my students to learn in groups
Q14 Using an IWB helps students to learn concepts easier

* This negative statement was reverse-coded

**Motivational effects of IWBs**

Teachers’ general attitudes and opinions related to IWB use were examined for motivational effects of IWBs in terms of either teachers or students (Table 3). Cronbach’s alpha coefficient for this part of the questionnaire is 0.89.

<table>
<thead>
<tr>
<th>Statements</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Disagree</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>In terms of teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15. I enjoy teaching with an IWB</td>
<td>172</td>
<td>3.97</td>
<td>.936</td>
<td>7.0</td>
<td>70.3</td>
</tr>
<tr>
<td>Q16. Because of using an IWB, I feel myself more prepared for instruction</td>
<td>162</td>
<td>3.73</td>
<td>1.002</td>
<td>11.7</td>
<td>67.9</td>
</tr>
<tr>
<td>Q17. I notice my IWB skills are improving day by day</td>
<td>162</td>
<td>3.90</td>
<td>.861</td>
<td>7.4</td>
<td>75.9</td>
</tr>
<tr>
<td>Q18. Learning how to use an IWB is essential to me</td>
<td>160</td>
<td>4.06</td>
<td>.856</td>
<td>5.6</td>
<td>81.3</td>
</tr>
<tr>
<td>Q19. IWB makes my courses more enjoyable</td>
<td>161</td>
<td>3.83</td>
<td>.875</td>
<td>7.5</td>
<td>67.1</td>
</tr>
<tr>
<td>In terms of students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q20. Using IWB increases my students’ interest in class</td>
<td>161</td>
<td>3.98</td>
<td>.877</td>
<td>4.3</td>
<td>75.2</td>
</tr>
<tr>
<td>Q21. My students look forward to my using an IWB in class</td>
<td>157</td>
<td>3.08</td>
<td>1.074</td>
<td>29.9</td>
<td>31.8</td>
</tr>
<tr>
<td>Q22. My students focus on my lessons more when I use an IWB</td>
<td>161</td>
<td>3.74</td>
<td>.925</td>
<td>8.7</td>
<td>64.0</td>
</tr>
<tr>
<td>Q23. IWB increases my students’ motivation towards the course</td>
<td>160</td>
<td>3.84</td>
<td>.843</td>
<td>6.3</td>
<td>70.0</td>
</tr>
</tbody>
</table>

**Usability of IWBs**

In order to examine the usability of IWBs, teachers were asked to respond to three statements (Table 4). Cronbach’s alpha coefficient for this group of questions is 0.60.

<table>
<thead>
<tr>
<th>Statements</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Disagree</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability of IWBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q24. IWB can be used in all kinds of courses 172 3.90 1.080 12.2 64.5
Q25. My course content is not suitable with using an IWB* 170 3.79 1.067 14.7 67.1
Q26. IWB can be used with various instructional methods and techniques 157 3.95 .830 3.8 74.5

* This negative statement was reverse-coded

Relationship between Teachers’ IWB usage and perceptions

Chi-square tests were administered in order to examine whether there are any differences between teachers’ perceptions in terms of the frequency and duration of their IWB use. Comparisons of individual differences in terms of the IWB use frequencies and also duration of IWB use on a weekly basis and perceptions about IWB use were analyzed by means of the chi-square test of independence. The results of the tests and effect size values (Cramer’s V) for each significant relation are presented in Table 5.

Table 5. Chi-square test results for IWB use and perceptions of teachers

<table>
<thead>
<tr>
<th>Items</th>
<th>Q4*</th>
<th>Q7*</th>
<th>Q11</th>
<th>Q17</th>
<th>Q24</th>
<th>Q28</th>
<th>Q39</th>
<th>Q52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of IWB use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(sometimes,</td>
<td>(\chi^2) (df) 18.92(8)</td>
<td>26.35(8)</td>
<td>18.13(8)</td>
<td>14.37(6)</td>
<td>18.66(8)</td>
<td>16.80(8)</td>
<td>15.58(8)</td>
<td>32.12(6)</td>
</tr>
<tr>
<td>(frequently,</td>
<td>Cramer’s V .235</td>
<td>.277</td>
<td>.232</td>
<td>.210</td>
<td>.240</td>
<td>.228</td>
<td>.221</td>
<td>.315</td>
</tr>
<tr>
<td>always)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of weekly IWB use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(up to 7 hours,</td>
<td>Items Q3</td>
<td>Q4</td>
<td>Q7</td>
<td>Q27</td>
<td>Q28</td>
<td>Q33</td>
<td>Q39</td>
<td>Q52</td>
</tr>
<tr>
<td>more than 7 hours)</td>
<td>(\chi^2) (df) 10.24(4)</td>
<td>10.07(4)</td>
<td>12.07(4)</td>
<td>8.10(3)</td>
<td>9.92(4)</td>
<td>10.80(4)</td>
<td>13.06(4)</td>
<td>11.32(3)</td>
</tr>
<tr>
<td></td>
<td>Cramer’s V .244</td>
<td>.242</td>
<td>.265</td>
<td>.224</td>
<td>.248</td>
<td>.262</td>
<td>.287</td>
<td>.264</td>
</tr>
</tbody>
</table>

As shown in Table 5, both frequency and duration of the IWB use moderately effect (Cramer’s V between 0.2 and 0.4) teachers’ perceptions in terms of the items presented above. In addition, we found significant results for three items (Q28, Q39, and Q52) in terms of both frequency and duration of the IWB use.

Discussion

Teachers’ IWB use and IWB training characteristics were somewhat strong. Since all participants were selected based on their prior experience with IWB use, it is not a surprise that a high percentage (79.3%) reported that they had ‘frequently’ or ‘always’ used IWBs in their courses. However, only half of them defined their IWB competencies above average. Although these findings give the sense about the high quantity of IWB use specifically for this sample, teachers were questioned on how they used IWBs in their classes and also what IWB features they used. The results regarding preferred features of IWBs were in parallel with the results of a previous study conducted by Türel (2011), which examined students’ perceptions about IWB use in Turkey. While teachers reported they used a wide range of IWB features, it is clear that several features including Internet and hyperlinks are the least preferred. For example, Beauchamp and Parkinson (2005) defined using hyperlinks as one of the highest levels of IWB skills according to the IWB progression levels. The data (see Figure 1) depicts the paucity of synergistic (professional) users based on the less utilized features and also the majority of initiate and advanced users based on Beauchamp’s (2004) framework, was consistent with the teachers’ self-reported competencies (see Table 1).

The main sources of teachers’ IWB training showed the majority of respondents (67%) had joined an IWB training session organized either by the company that supplies IWBs or by an educational institution. Looking at learning resources, almost half of the respondents (44%) marked ‘colleagues’ as the main resource for their support and while about a quarter of respondents reported that they learned to use IWBs by themselves. Shenton and Pagett (2007), who observed IWB users in the UK schools, suggest that their teacher-participants gained skills to use IWBs working by themselves or with a group of teachers. As such, these two sources should be considered as essential supports for teachers’ IWB use as well as for informal IWB training.

Teachers’ ratings for the three IWB-training themes surprisingly indicated that most of them do not need IWB training related to finding and designing instructional materials for the IWB although only 22 teachers received
training regarding this topic. This result may stem from teachers’ learning IWB on their own. However, one third of all participants still reported that they need IWB training in each one of the three areas: IWB technical knowledge and skills; teaching methods related to IWB; and designing IWB activities. This finding is broadly in line with the results of several studies (Glover & Miller, 2001; Smith et al., 2005; Somyürek, Atasoy, & Özdemir, 2009).

Effects of Interactive Whiteboards

Teachers’ perceptions and attitudes about the IWB use were studied looking at three main themes: instructional effects, motivational effects, and usability. The first theme is related with the effects of IWBs on teaching and learning processes and also addresses the advantages of IWBs. Similar to the results of previous IWB studies (Beeland, 2002; Moss et al., 2007; Erduran & Tataroğlu, 2009; Mathews-Aydinli & Elaziz, 2010; Saltan, Arslan, & Gök, 2010), it is apparent that teachers have positive perceptions (3.79/5.0) about the use of IWBs in general.

Focusing on IWB advantages, the seventh item has the highest mean score (M = 4.41, SD = .701). This item basically refers to the advantages relating to the visualization of course content. Teachers can design and use visually attractive materials compatible with an IWB; moreover, they can enhance their presentations, before or during instruction, with visual effects including highlighting, coloring, drawing, zooming, or can import visual objects from other sources (e.g., web-pages, Paint) via using screenshot or copy-paste feature (Türel, 2010). Wall, Higgins, and Smith (2005) suggest that such presentations help teachers to draw student attention to course content and also facilitate student retention of what they learned and facilitate student understanding of concepts (Levy, 2002). Those benefits are essential elements for students’ learning; which may explain why teachers overwhelmingly agreed (77%) that they believed that using IWBs helps their students’ learning (M = 4.16, SD = .904, see Q10 in Table 2).

The agreement level for the fifth question, ‘IWB facilitates discussions on the content in class’ (M = 3.30, SD = .954) reveals teachers’ neutrality about IWB effect for class discussions. BECTA (2006) suggest that an essential IWB strategy would include using IWBs for initiating discussions about the course content. Teachers can use an IWB to share content for a class or a small group discussion. An IWB can be used for students to share their ideas in a discussion setting. Teachers are expected to improve their skills in terms of effectively using IWB strategies and, to promote their instructional activities based on the promises of IWBs (Beauchamp & Parkinson, 2005).

Teachers were asked the question: ‘The way I give instruction has changed since I began to use an IWB (Q8)’. Almost half of the teachers agreed or strongly agreed (49.4%) to this statement (M = 3.45, SD = 1.011), which can mean that some level of pedagogical change may have occurred due to IWB technologies. Two items in the questionnaire are related to managerial issues of instruction: time management (Q1) and classroom management (Q3). According to the findings, most teachers believe that IWB provides time efficiency during instruction (M = 4.27, SD = .943). Likewise, researchers such as Levy (2002) and Tozcu (2008) suggest that using an IWB reduces the time spent recreating instructional materials and content since teachers have an electronic copy from the IWB were as they do not have an electronic copy with traditional boards. In addition, teachers can easily interact and communicate with students using an IWB and keep students engaged during a lesson. This is regarded as a major benefit of IWBs in terms of classroom management (BECTA, 2003). However, teachers who only present course content with an IWB in the same manner as a data projector may not give students an opportunity to sufficiently use it during instruction. Such kinds of teacher-centered practices may lead to a decrease in students’ attention and motivation. Aligned with constructivist perspectives, teachers can encourage students to actively participate in the learning process by working on the IWB individually or in groups (Smith et al., 2005). However, the question, ‘There is no time for my students to get around to using an IWB (Q6)’ has the lowest mean score (M = 2.99, SD = 1.098). This indicates the possibility of a predominant teacher-centered modality in the classroom. However for effective teaching and learning, current studies (BECTA, 2003; Smith et al., 2005) maintain the importance of a student-centered modality using IWBs.

On the other hand, most teachers agreed that using an IWB is motivating, engaging, and enjoyable for both teachers and students. This finding is parallel with other studies (Bell, 2002; BECTA, 2003; Smith et al., 2005; Mathews-Aydinli & Elaziz, 2010). The lowest mean for motivational effects theme was the statement: ‘My students look forward to my using an IWB in class (Q21)’ (M = 3.08, SD = 1.074). Teachers’ moderate responses to this question may result from students’ willingness to use the IWB themselves or could be attributed to the decrease of students’ initial enthusiasms towards IWB use more formally referred to as the novelty effect (Levy, 2002; BECTA, 2003).
Another key finding is that most teachers (75.9%) strongly agree that they are aware of the continuous improvement of their IWB skills (Q17) \( (M = 3.90, SD = .861) \). This finding indicates that one of the key sources for teachers IWB skill development is coming from their own experience as shown in Figure 2. Furthermore, teachers agreed on the importance of learning to use an IWB (Q18) \( (M = 4.06, SD = .856) \) by reporting how valuable the IWB is for their instruction.

For the usability theme, teachers were asked to indicate whether IWBs can be used in different contexts and ways. It is evident that teachers’ perceptions regarding to usability of IWBs in any kind of course and course content are positive (Q24) \( (M = 3.90, SD = 1.080) \). Three quarters of the teachers agreed that IWBs can be used with various instructional methods and techniques (Q26) \( (M = 3.95, SD = .830) \). These findings suggest that IWBs are not tied to a specific context. Teachers who participated in this study have positive attitudes about the usefulness and usability of IWBs. These attitudes are essential indicators in terms of the acceptance and the prediction of effective use of this technology, as outlined by Davis’s (1989) model.

Similarly, the frequency and the duration of technology use are other essential indicators for the acceptance of technology. Several significant differences were found in this study regarding the frequency and duration of teachers’ IWB use. Expectedly in this study, teachers who frequently used an IWB were more likely to have a higher level of IWB competency and more positive perceptions towards an IWB use as suggested by Moss et al. (2007). As Glover et al. (2007) stated, ‘teachers need time to develop their technological fluency, apply pedagogic principles to the available materials or to the development of materials, and then to incorporate the IWB seamlessly into their teaching’ (p. 17). Therefore, encouraging teachers to use an IWB more frequently may help them 1) to effectively integrate the IWB in their instructions, 2) to have more positive attitudes towards IWBs, 3) to accept this technology as an effective and a helpful instructional tool, and also 4) to cope with the emerging issues of IWB use as competent users. This finding also confirms the importance of teachers’ individual efforts to achieve higher-level IWB skills and knowledge as emphasized in the findings regarding the source of IWB skills and knowledge.

Concerning teachers’ effective use of any technology, several issues are associated with each other such as acquiring appropriate skills and knowledge, perceived efficiency, and usage frequency of the technology. Even after a comprehensive IWB training session, teachers who do not sufficiently use an IWB and do not practice what they have learned may have lost their initial IWB skills and knowledge as well as their confidence over time (Slay, Siebörger, & Hodgkinson-Williams, 2008). Hall and Higgins (2005) emphasize that teachers need continuous training sessions to improve and also maintain such skills. It is clear that teachers in this study need training particularly on using effective instructional strategies for IWB-assisted courses in order to transform their pedagogy into more student-centered, social and interactive learning. However, there are two existing problems: 1) one time training sessions provided by the representative of IWB supplier are superficial, and 2) schools do not have the time and budget to provide regular training sessions. As such, teachers should be supported to continuously use IWBs in their classrooms by working with their peers in order to improve their IWB skills and knowledge as suggested by Shenton & Pagett (2007). Furthermore, effective IWB implementations shared by groups of teachers may increase the awareness of teachers in terms of effective IWB strategies as well as their positive attitudes towards using the IWB in their courses. Teachers can find just-in-time solutions to their IWB problems. These kinds of practice help teachers to quickly overcome the wow barrier that Beauchamp and Parkinson (2005) stressed as a breakpoint to achieve a profound impact of the IWB on instruction.

**Limitations**

This study, which represents a snapshot of IWB use, has several limitations that may provide guidance for future research. For example, a qualitative analysis would be helpful for the examination of the underlying reasons of significant differences emerging between the teachers who most and least frequently use IWBs in their classes. On the other hand, the questionnaire developed specifically for this study considering the existing literature and associated theories and models. This instrument may be exposed to confirmatory factor analysis to examine the fitness of themes with various fit-indexes (Anderson & Gerbing, 1988). Finally, similar research may be conducted by considering additional IWB factors such as issues of IWB-assisted courses and effects of receiving IWB trainings on the IWB use.
Conclusion

This study provides a solid example of IWB integration and IWB effects on the teaching and learning process, in a rapidly developing country. It should be noted that this work neither reflects the status quo of IWB use in general, nor investigates the challenges and technical issues of IWBs in Turkey. Rather, it attempts to uncover the more realistic effects of using IWBs for teaching and learning by recruiting active IWB users from various fields.

The findings from this study demonstrate the key characteristics and strategic requirements of effective IWB use based on the perceptions of teachers who were active IWB users. For better understanding and interpretation of teachers’ perceptions, it is important to represent their background, as shown in this study, regarding IWB use including the frequency of IWB use, IWB competency, sources of IWB skills, and demographics. In general, participants were satisfied with the IWB use and they accepted IWBs as a powerful and practical technology that facilitates teachers’ instructions as well as students’ learning and motivation. However, findings indicate that teachers were not able to design a social constructivist environment where students could be involved in active and collaborative learning process using IWBs. Interestingly, most teachers believed that IWBs provided time efficiency for their instruction; however, a majority of them admitted that they could not find enough time to let their students use IWBs.

Results of the study also indicate a moderate correlation \((p < .05)\) in the relationships between both the frequency of IWB use and perceptions about IWBs as well as the duration of IWB use and perceptions about IWBs. These variables are regarded as the key factors for effective use of technology. In addition, most teachers confirmed that their IWB skills were improved as they used the IWBs and stated that they learned IWBs mainly from their colleagues. Therefore, it seems critical to support teacher IWB collaborations.

It is expected that the findings of this study may help teachers and researchers who are interested in effective IWB use and also administrators who are responsible for integration of ICT or organizing IWB training sessions. IWBs have the potential to engage students’ in various activities thereby supporting their learning and development. However if we are to expect students to improve their learning in the classroom, teachers need to develop their technology skills and positive attitudes through continued collaborative training and practice.

Acknowledgements

The authors would like to acknowledge the fellowship provided to the first author by The Scientific and Technological Research Council of Turkey (TUBITAK).

References


