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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.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; Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Institut Angewandte Informationstechnik, Germany.

Editorial Assistant

Barbara Adamski, Athabasca University, Canada.

Associate editors

Nian-Shing Chen, National Sun Yat-sen University, Taiwan; Alexandre I. Cristea, Technical University Eindhoven, The Netherlands; John Eklund, Access Australia Co-operative Multimedia Centre, Australia; Vladimir A Fomichev, E. Tsiolkovsky Russian State Tech Univ, Russia; Olga S Fomicheva, Studio "Culture, Ecology, and Foreign Languages", Russia; Piet Kammers, University of Twente, The Netherlands; Chai-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; Luis Anido-Rifon, University of Vigo, Spain; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Darina Dicheva, Winston-Salem State University, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Roger Hartley, Leeds University, UK; Harald Haugen, Hogskolen Stord/Haugesund, Norway; J R Isaac, National Institute of Information Technology, India; Mohamed Jenni, University of Tunis, Tunisia; 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; Ruddy Lelouche, Universite Laval, Canada; David McConnell, Lancaster University, UK; Rory McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Thomas C. Reeves, University of the Georgia, USA; Gilly Salmon, University of Leicester, United Kingdom; 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; Brian K. Smith, Pennsylvania State University, USA; J. Michael Spector, Florida State University, USA; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Stephen J.H. Yang, National Central University, Taiwan.

Assistant Editors

Sheng-Wen Hsieh, Far East University, Taiwan; Taiyu Lin, Massey University, New Zealand; Kathleen Luchini, University of Michigan, USA; Dorota Mularczyk, Independent Researcher & Web Designer; Carmen Padrón Nápoles, Universidad Carlos III de Madrid, Spain; Ali Fawaz Shareef, Massey University, New Zealand; Jarkko Suonen, University of Joensuu, Finland.

Executive peer-reviewers

http://www.ifets.info/

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.

ISSN 1436-4522 (online) and 1176-3647 (print). © International Forum of Educational Technology & Society (IFETS). The authors and the forum jointly retain the copyright of the articles. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the full citation on the first page. Copyrights for components of this work owned by others than IFETS must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from the editors at kinshuk@ieee.org.
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 a 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, they can also be sent via email to (Subject: Submission for Educational Technology & Society journal): kinsbuk@ieee.org. In the email, please state clearly that the manuscript is original material that has not been published, and is not being considered for publication elsewhere.
Table of contents

Full length articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Effects of Metacognitive Knowledge on the Pre-service Teachers’ Participation in the Asynchronous Online Forum</td>
<td>Abdullah Topcu and Behiye Ubuz</td>
<td>1-12</td>
</tr>
<tr>
<td>Applying Web-Enabled Problem-Based Learning and Self-Regulated Learning to Add Value to Computing Education in Taiwan’s Vocational Schools</td>
<td>Tsang-Hsiung Lee, Pei-Di Shen and Chia-Wen Tsai</td>
<td>13-25</td>
</tr>
<tr>
<td>Who is responsible for E-Learning Success in Higher Education? A Stakeholders' Analysis</td>
<td>Nicole Wagner, Khaled Hassanein and Milena Head</td>
<td>26-36</td>
</tr>
<tr>
<td>A Survey on ICT Usage and the Perceptions of Social Studies Teachers in Turkey</td>
<td>Yasemin Gulbahar and Ismail Guven</td>
<td>37-51</td>
</tr>
<tr>
<td>A Study of Uses of ICT in Primary Education through Four Winning School Cases in the Taiwan Schools Cyberfair</td>
<td>Shelley S.-C. Young and Hsin-Ho Ku</td>
<td>52-66</td>
</tr>
<tr>
<td>Experimental Evaluation of an Instructional Supporting Tool in Distance Learning</td>
<td>Athanasis Karoulis, Ioannis Stamelos and Lefteris Angelis</td>
<td>67-81</td>
</tr>
<tr>
<td>Beyond Sharing: Engaging Students in Cooperative and Competitive Active Learning</td>
<td>Gloria Yi-Ming Kao, Sunny S. J. Lin and Chuen-Tsai Sun</td>
<td>82-96</td>
</tr>
<tr>
<td>Testing Principles of Language Learning in a Cyber Face-to-Face Environment</td>
<td>Nian-Shing Chen and Yuping Wang</td>
<td>97-113</td>
</tr>
<tr>
<td>Effects of Text, Audio, and Graphic Aids in Multimedia Instruction for Vocabulary Learning</td>
<td>Daesung Kim and David A. Gilman</td>
<td>114-126</td>
</tr>
<tr>
<td>Training in spatial visualization: The effects of training method and gender</td>
<td>Ahmad Rafi, Khairul Anuar Samsudin and Che Soh Said</td>
<td>127-140</td>
</tr>
<tr>
<td>Competence Description for Personal Recommendations: The importance of identifying the complexity of learning and performance situations</td>
<td>Frans J. Prins, Rob J. Nadolski, Adriana J. Berlanga, Hendrik Drachsler, Hans G.K. Hummel and Rob Koper</td>
<td>141-152</td>
</tr>
<tr>
<td>Personalized Intelligent Mobile Learning System for Supporting Effective English Learning</td>
<td>Chih-Ming Chen and Shih-Hsun Hsu</td>
<td>153-180</td>
</tr>
<tr>
<td>Effects of the Asynchronous Web-Based Course: Preservice Teachers’ Achievement, Metacognition, and Attitudes towards the Course</td>
<td>Abdullah Topcu and Behiye Ubuz</td>
<td>181-197</td>
</tr>
<tr>
<td>Building a Model Explaining the Social Nature of Online Learning</td>
<td>I-Chun Tsai, Bosung Kim, Pei-Ju Liu, Sean P. Goggins, Christiana Kumalasari and James M. Laflaye</td>
<td>198-215</td>
</tr>
<tr>
<td>Considering Students’ Perceptions: The Distance Education Student Satisfaction Model</td>
<td>Ismail Sahin and Mack Shelley</td>
<td>216-223</td>
</tr>
<tr>
<td>A study of the relationship between student social networks and sense of community</td>
<td>Shane Dawson</td>
<td>224-238</td>
</tr>
<tr>
<td>Modeling and Intervening Across Time in Scientific Inquiry Exploratory Learning Environment</td>
<td>Choo-Yee Ting, Somnak Phon-Ammatassuk and Yan-Kuan Chong</td>
<td>239-258</td>
</tr>
</tbody>
</table>
A Practical Computer Adaptive Testing Model for Small-Scale Scenarios
   Yu-Hui Tao, Yu-Lung Wu and Hsin-Yi Chang
   259-274

Comparison of Dimension Reduction Methods for Automated Essay Grading
   Tuomo Kakkonen, Niko Myller, Erkki Sutinen and Jari Timonen
   275-288

**Book review(s)**

Blended learning: Using technology in and beyond the language classroom
   Reviewer: Jesús García Laborda
   289-291

Semantic Web and Education
   Reviewer: Fuhua Oscar Lin
   292-293

**Website review(s)**

National Library of Virtual Manipulatives
   Reviewer: Hosin Shirvani
   294-295
The Effects of Metacognitive Knowledge on the Pre-service Teachers’ Participation in the Asynchronous Online Forum

Abdullah Topcu
Measurement and Evaluation Office Army School of Language (serves to NATO personnel), Istanbul, Turkey
Tel: +90 216 518 96 81 // abdullah.topcu@boun.edu.tr

Behiye Ubuz
Faculty of Education, Middle East Technical University, Ankara, Turkey // Tel: +90 312 210 40 62 // ubuz@metu.edu.tr

ABSTRACT
This paper researches the effects that students’ metacognitive knowledge has on their participation in online forum discussions, which form part of a web-based asynchronous course based on a constructivist instructional approach. Metacognitive knowledge increases learners’ ability to be independent learners, which is an indispensable characteristic of the distant learner. The study was carried out with 32 third-grade pre-service teachers. Each message in the forum discussions was analyzed in terms of interaction types identified by McKinnon (2000) and also scored using a grading rubric developed by the researchers. The metacognitive knowledge of the pre-service teachers was measured by the component of the General Metacognition Questionnaire. Sixty-seven percent of the pre-service teachers were at the high or medium-to-high metacognitive knowledge level and mostly sent messages having “example to idea,” “clarification and elaboration,” or “idea to example” type interactions. Pre-service teachers who exhibited low metacognitive knowledge, however, mostly forwarded messages having “acknowledgments,” “unsubstantiated judgment,” or “thoughtful query” type interactions. Moreover, metacognitive knowledge of the pre-service teachers uniquely explained 21.4% of the variance in the online participation score. We concluded by outlining some implications metacognitive knowledge has on forum discussions in relation to the constructivist approach.

Keywords
Metacognitive knowledge, Constructivist approach, Online participation, Interaction

Introduction
Over the last decade, web-based asynchronous learning environments (WALE) have become widely used for instructional purposes (Holmes, 2005). A constructivist instructional approach is the main stream in WALE (Jonassen, 2000; Tam, 2000), transforming it into active learning environments. A crucial transporter of active participation in the constructivist WALE is messages forwarded in the forum discussions. The discussion forum is an example of an asynchronous/synchronous tool in a WALE used for constructing and managing knowledge collaboratively through posing certain types of engaging activities such as facilitative/guiding questions and examples of real cases (e.g., videos, episode dialogues), amongst others. Frequency, depth, and cohesiveness of the messages and interactivity among them heavily rely on the surrounding context of learners (Yang, 2006). Context is a process whereby a person consciously or unconsciously compares an external environment with acquired personal experiences/knowledge (both of which may contain task, physical, social, and temporal dimensions) in order to form goals for undertaking concise actions, possibly with other people and/or objects (Bradley, 2005). In WALE, students’ metacognition (Ahitikari & Eronen, 2004; Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Moore, 2002; Tsai, 2004) is one of the most important personal variables that affect the engagement in the learning task, a deeper level of understanding, the development of higher-order thinking, the information-seeking behaviors, the processing of information gathered in online environments, etc. For students to produce efficient participation and deeper levels of thought in forum discussions, metacognitive strategies that regulate self-awareness, self-control, and self-monitoring are necessary. Activities such as planning how to approach a given learning task, monitoring comprehension, evaluating progress toward the completion of a task, and knowledge of these activities are metacognitive in nature. It stands to reason that an important step in enhancing learning outcomes in WALE is to obtain clarity with regard to the influence of metacognition on how and how often learners bring themselves into the learning process in online forum discussions.
Metacognition can be defined as “one’s knowledge concerning one’s own cognitive processes” (Flavell, 1979). As a construct, it is based on information processing and general intelligence theories (Pintrich, Wolters, & Baxter, 2000). Brown, Campione, and Barclay (1979) defined it as the control processes in which active learners engage as they perform various cognitive activities. Initially, some researchers have proposed that the term metacognition be reserved for the construct of metacognitive knowledge and that the term not include metacognitive control and regulation activities (Paris & Winograd, 1990). But later, a general agreement revealed that metacognition can be divided into three different constructs (Pintrich, Wolters, & Baxter, 2000). They are (a) metacognitive knowledge (b) metacognitive judgments and monitoring, and (c) self-regulation and control of cognition. The relations of the metacognition with the autonomy and interaction push the attention on the roles, developments, and effects of the students’ metacognitive abilities in a web-based learning environment. Moore (2002) claims that before students can make sense of the vast store of information in a web-based learning environment, they have to develop certain advanced metacognitive skills and web literacy. In fact, web literacy also includes the metacognitive awareness of oneself as a user of the environment (Ahitikari & Eronen, 2004). Until now, researchers (Erlich, Erlich-Philip, & Gal-Ezer, 2005; Tobias & Everson, 2000; Weinberger & Fischer, 2006) generally have studied the components of the metacognition in the online learning environment for different reasons. The most important reason is to find specific, applicable, and effective approaches that facilitate producing learning outcomes (Shraw, 2000).

Therefore, this study attempted to investigate the relationship between metacognitive knowledge (MK), one of the components of the metacognition, and participation in asynchronous online discussions. MK refers to acquired knowledge about cognitive processes and knowledge that can be used to control cognitive processes. Flavell (1979) proposed that MK includes knowledge of self (e.g., knowing that simulation is better at memory tasks than problem-solving tasks or knowing that one has to pay close attention to something in order to learn it), tasks (e.g., knowing that tasks vary in their level of difficulty or knowing that it will take more time for a person to read and comprehend a science text than it would for a person to read and comprehend a novel), and strategy variables or factors that can influence cognition (e.g., knowing that rehearsal can help comprehension of the main idea in the messages).

The question posed in this paper was: Does pre-service teachers’ metacognitive knowledge affect their interaction and message levels as well as the discussion scores within the forum discussions of the web-based asynchronous course “science and math teaching method,” which was developed based on the constructivist instructional approach? From this follow three related questions: (1) How are the types of online interactions related to the pre-service teachers’ level of MK? (2) How are the message levels (initial posting-1st level, response to initiating post-2nd level, so on) related to the MK level of the pre-service teachers? (3) Is there a relationship between MK level and pre-service teachers’ discussion scores? In the light of these questions, the goals of this paper are two-fold. The immediate goal is to provide different perspectives for the instructor or moderator to be able to create interactive, cohesive and effective online forum discussions by taking into account students’ MK. A longer-term goal is to create an awareness of the importance of producing various MK strategies which can be embedded into instructional methods for online delivery via WALE.

**Metacognitive knowledge and asynchronous online discussions**

Asynchronous online discussions are increasingly integrated into the online as well as traditional face-to-face courses since they open unprecedented opportunities for educational interactivity. They provide a virtual and conversational learning environment in which students are likely to learn as much from one another as from course materials or lectures. Online discussions allow students and teachers alike to share their own perspectives and analyze others’ views, simulating a real class discussion in a face-to-face situation. By using various sources such as the WWW, books, and experiences, students can research and/or think about the topic, build their arguments in anticipation of opposing arguments, and plan rebuttals with reasoning and evidence. All these cognitive processes engaged in learning involve knowledge about cognitive processes and knowledge that can be used to control cognitive processes. In this regard, MK can influence how individuals explore, identify, and monitor their thinking and learning in an online forum discussion (Gilbert & Dabbagh, 2005; Hill & Hannafin, 1997), as well as the depth of their learning, which ensures accurate knowledge construction through the online discussions (Hew & Cheung, 2003; Srinivasan & Crooks, 2005).

MK influences learners to effectively participate in online discussions, to comprehend main ideas in online discussions, to construct connections between previous and new knowledge in message sequences, and to evaluate
available resources. Learners who effectively use MK are largely strategically inclined and might have a cognitive platform to identify relevant content and applicable services in the right place at the right time depending on their context. Such learners have the ability to reorientate themselves spatially with greater prowess and identify what learning resources and services are available to them. They furthermore identify fellow learners who can act as learning collaborators that match their needs (Zhang, Jin, & Lin, 2005).

In asynchronous communication, the recording of all students’ messages in the courseware allows students to review the pathways navigated and evaluate what they have done in the process of discussing ideas (Hurme & Jarvela, 2005). In addition, mutual message exchange between participants can take place at any time, so participants do not have to wait for explicit turn-taking (Schwan Straub, & Hesse, 2002). Participants, as a basic precondition, have to read the messages and ask questions, make comments, and provide answers. Actively participating learners realise the entire spectrum of these activities, whereas passively participating learners restrict their activities to reading the messages. For active participation, producing a written contribution (in text-based communication), inserting it into a topic and comment structure, and addressing it to the desired addressees are necessary (Hron & Friedrich, 2003). Beyond this, learners have to decide when and how often they bring themselves into the learning process (e.g., once a week) and, usually depending on the number of messages, how long they will take part and which contributions they will read and answer (Schwan, Straub, & Hesse, 2002).

All of these cognitive processes engaged in active participation include knowledge of various learning strategies (e.g., rehearsal, elaboration, organizational). Students use these strategies to extract meaning from messages, text, or the WWW and to comprehend what they read in messages. In addition, knowledge of various metacognitive strategies (e.g., setting subgoals, asking themselves questions as they read a message, checking responses...) are useful to students in planning, monitoring, and regulating their thinking through asynchronous online discussions. Furthermore, knowledge about various strategies and knowledge of tasks — knowledge that different tasks can be more or less difficult and may require different cognitive strategies — are indispensable for active participation to reach meaningful learning in WALE (Gilbert & Dabbagh, 2005). For example, in producing a written contribution to the asynchronous online discussions, the emphasis is on the formulation of new ideas or structure. Lastly, students who know their own strengths and weaknesses can adjust their own cognition and thinking to be more adaptive to diverse tasks and, thus, facilitate their participation and learning (Erlich, Erlich-Philip, & Gal-Ezer, 2005). If, for example, a student realizes that she does not know very much about a particular topic, she might pay more attention to the topic while reading and use different strategies to make sure she understands the topic being studied. Therefore, MK might have significant roles for effective participation and, consequently, meaningful learning.

**Methods**

**Participants**

The sample consisted of third-year undergraduate science and mathematics pre-service teachers in the department of elementary education at a state university where the educational language is English. A total of 32 pre-service teachers (12 males and 20 females) participated in the study, with ages ranging from 21 to 24. Of the participants in the study, there were 14 science education and 18 mathematics education majors. Ninety-two percent of the pre-service teachers had previously completed at least one computer-related course, such as “computer-assisted instruction” and “introduction to computers.” About half of them surf the Internet/web frequently. None of them has experienced an online course since no one had enrolled for a web-based course before.

**Instruments**

*General Metacognition Questionnaire*

The items related to the MK in the General Metacognition Questionnaire (GMQ) were used to determine the participants’ level of the metacognitive knowledge. GMQ covers all the components of metacognition: metacognitive knowledge, metacognitive judgments and monitoring, and self-regulation and control of cognition.
Eight of the 30 items in GMQ are related to metacognitive knowledge. Examples of MK items are (a) “When I study, I practice saying to myself the important facts, principles, and concepts in my cognition,” and (b) “I begin to study material by asking myself what I know and what I do not know.” Pre-service teachers were asked to rate 30 statements by making judgments about previous learning conditions. They had to answer by marking a five-point Likert scale with anchorpoints noted as 1) “Not true for me,” 2) “Hardly true for me,” 3) “I don’t know,” 4) “Rather true for me,” and 5) “Very true for me.” The midpoint (3) was interpreted as “may be it is correct; may be it is not correct.” Previously, the authors used the GMQ within 221 pre-service teachers to investigate its psychometric properties. An overall review of validity data suggests the GMQ measures that it aims to measure. The obtained Cronbach’s alpha was .72.

Observation checklist for the constructivist approach

To determine whether the web-based asynchronous course, “science and mathematics teaching methods,” was conducted according to the constructivist approach or not, an observation checklist including 15 items that contain the criteria of the constructivist instructional approach was developed by considering the literature related to the constructivist approach (Jonassen, 2000; Tam, 2000). Items contained criteria related to the instructor (four items), students (seven items), and medium (four items). Examples of items from each criterion included (a) “Instructor acts as a coach, not a content provider,” (b) “Students focus on knowledge building rather than knowledge reproduction,” (c) “Course site provides links related to the course content to stimulate recall of prior learning,” respectively. There were six negative and nine positive statements in the scale, with five possible alternatives: (1) Strongly disagree, (2) Disagree, (3) Uncertain, (4) Agree, and (5) Strongly Agree. Each statement was graded as 1, 2, 3, 4, and 5. Negative statements were scored in reverse.

Web-based asynchronous course, “Science and mathematics teaching methods” (WBAC)

The web-based course “science and mathematics teaching methods” includes ten chapters and relies on asynchronous forum discussions. The chapters are offered in the following sequence: (1) Introduction to the course and the effective teacher, (2) Questioning in the classroom, (3) Introduction to methods and direct instruction, (4) Discovery teaching, (5) Problem solving, (6) Discussion, (7) Cooperative learning, (8) Computer-assisted instruction, (9) Project-based learning, and (10) Drama. The co-researcher of this study has been teaching the course for eight years and has verified the course content.

Each chapter lasted one week, except the first one. The first chapter continued one and half weeks as the pre-service teachers got accustomed to the WBAC environment and its tools. The flow of the week for each chapter except the first one is outlined in Figure 1. There were three periods for each week. First-period discussion was about the questions centered on the concepts of the topic, whereas second-period discussion was about the questions on the ill-structured cases. The ill-structured cases were presented in the form of the video clips for chapters 1, 5, 6, 7, and 8 and the episodes narrating the real classroom discourse for the rest of the chapters. In the third period, there was no discussion, only evaluation. For each week, the instructor initiated three or four threads (main topics of the chapters) to be discussed. For example, for the “Introduction to the course and effective teacher” chapter, four threads were initiated: (1) Some terms used in the course, (2) Effective teacher, (3) Teacher’s level of knowledge of subject matter, (4) Psychological characteristics of effective teachers. Under each thread, three or four main facilitative/guiding questions were directed in the first two periods. These questions were prepared prior to the administration of the WBAC by considering the content of chapters and the principles of the constructivist instructional approach explained in the literature (Jonassen, 2000; Willis, 1995). For the second thread, “Effective teacher,” in the first period, the following question was posed: “Recall both effective and ineffective teachers you may have had. To what extent did they seem to differ with respect to the knowledge of how to teach the subject, or knowledge of how people learn?”

The forum discussions started with questions such as the one above. The pre-service teachers were given the opportunity to search the Internet and e-library in the course site, mindful of the topics given in the outline of each chapter and the questions posted in the forum. In addition, the pre-service teachers carried out their searches by using the books cited in the syllabus. The first period continued between Monday at 10 a.m. and Thursday 10 a.m. During that period on Wednesday at 10 a.m., two of the pre-service teachers were tasked with summarizing the forum.
discussions. The due date of the constructed summaries of the first period was Friday 10 p.m. In the second period, the ill-structured cases (video clip or dialog in the episode) and the main facilitative/guiding questions that were related to them were posted to the forum on Thursday at 10 a.m. As a result, the second period of the discussion began and continued until Friday 10 p.m. Pre-service teachers participated in discussion after watching the video clip or reading the dialog in the episodes. For example, in chapter one, a video clip of a math class was presented. One of the forwarded questions related to the video clip in the third thread, “Teacher’s level of subject matter knowledge” was as follows:

Is the instruction for this period effective if you consider the “teacher task orientations” and “student engagement”? Why or why not?”

One of the pre-service teachers was also tasked with summarizing the forum discussions carried out in the second period on Friday at 10 a.m. The summaries were due Saturday at 10 p.m. In the last period, the lecture notes were written by a pre-service teacher. He or she was chosen based on discussions during the first two periods. The announcement of this task was also made on Friday at 10 a.m. The pre-service teachers returned the lecture notes to the instructor on Sunday at 10 p.m. The instructor published the lecture notes of the pre-service teachers on the course site on Monday at 10 p.m., after checking the content for lack of important knowledge or faults according to main assumptions of the topics. All of the pre-service teachers took the duty of either summarizing forum discussions or preparing lecture notes at least once.

Each pre-service teacher had to participate at least three times each week in the discussions. In this way, a discussion environment was formed without making the participation into a burden for the pre-service teachers. Meanwhile, this rule caused an increase in interactivity. Each message of the pre-service teachers in the forum discussions was assessed in terms of interaction types of the coding technique developed by McKinnon (2000). This coding technique was chosen among others (Henri, 1992; Gunawardena, Lowe, & Anderson, 1997; Garrison, 1992; Newman, Webb, & Cochrane, 1996), because it focuses only on the interactivity. Table 1 illustrates the coding technique used in the
present study. Meanwhile, a grading rubric developed by the researchers was used to score the pre-service teachers’ messages, thereby determining the quality of their participation. The rubric is shown in Table 2. The discussion scores of the pre-service teachers related to each week were announced on Sunday.

Table 1. Coding technique of messages

<table>
<thead>
<tr>
<th>Specific Interaction</th>
<th>Code Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgement of Opinions</td>
<td>A</td>
</tr>
<tr>
<td>Question (thoughtful query)</td>
<td>Q</td>
</tr>
<tr>
<td>Compare (similarity, analogy)</td>
<td>CM</td>
</tr>
<tr>
<td>Contrast (distinction, discrimination)</td>
<td>CN</td>
</tr>
<tr>
<td>Evaluation (unsubstantiated judgment, value)</td>
<td>EV</td>
</tr>
<tr>
<td>Idea to example (deduction, analogy)</td>
<td>I2E</td>
</tr>
<tr>
<td>Example to idea (induction, conclusion)</td>
<td>E2I</td>
</tr>
<tr>
<td>Clarification, elaboration (reiterating, building on a point)</td>
<td>CLE</td>
</tr>
<tr>
<td>Cause and effect (inference, consequence)</td>
<td>CE</td>
</tr>
<tr>
<td>Off-topic / faulty reasoning (entry inappropriate)</td>
<td>OT</td>
</tr>
</tbody>
</table>

Table 2. Grading rubric of messages

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putting comprehensive title to the message that is consistent with its content.</td>
<td>2</td>
</tr>
<tr>
<td>Introducing the content truly and cohesively</td>
<td>2</td>
</tr>
<tr>
<td>Explaining the content based on investigation or examples or peers’ ideas</td>
<td>3</td>
</tr>
<tr>
<td>Summary of the message content</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
</tr>
</tbody>
</table>

Procedure

The “methods of science and math teaching” course is a compulsory three-credit second-semester course for the third-year undergraduate mathematics and science education pre-service teachers in the Department of Elementary Education at a state university. Although offered online, there were three face-to-face classroom meetings throughout the semester. At the first meeting (at the beginning of the semester), students were introduced to the course. This involved a three-hour demonstration of the online forum discussion tools in the computer laboratory. The second meeting was for the midterm exam, and the last one was for the final exam. The instructor was the first author of this study.

Using an observation checklist, two raters independently verified whether WBAC was constructivist or not. The inter-rater reliability was calculated, and the correlation was found as significant, \( r = .90, p < .05 \). In addition, the first researcher rated six different chapters’ course flows by using the same observation checklist. Means of criteria containing positive statements were significantly greater than 3.80. Only the mean of positive criteria, “students’ focus on knowledge-building rather than knowledge reproduction,” was moderately higher than 3 (3.2). Means of criteria containing negative statements were less than 2. According to these results, it can be assumed that WBAC was conducted according to the constructivist approach.

Each message forwarded by the students in the forum discussion was analyzed with respect to the McKinnon’s Coding Technique to identify the interaction types. Each message took one interaction code. The primary test of objectivity in content analysis is inter-rater reliability. The interaction types agreement was 0.76 (Cohen’s kappa) between the above two raters. This value is accepted as fair-to-good agreement beyond chance (Cappozoli, McSweeney, & Sinha, 1999). A sample of interaction types is illustrated in Table 3. Further, each message was scored using the grading rubric given in Table 2. Inter-rater reliability for the discussion scores conducted by the above two raters was found as significant, \( r = .91, p < .05 \). For the content analysis data that did not reach the raters’
agreed-upon categorization, the raters reviewed the transcripts again and discussed them on a case-by-case basis, and then determined the final interaction types and scoring. Both raters carried out the classifications separately and negotiated their differences of opinion in order to reach consensus.

The GMQ used in this study was administered as a pretest by email at the first week of the course and pre-service teachers were required to return the questionnaire by an email attachment to the instructor in the same week. A formal questionnaire was used in the study to assess students’ metacognitive knowledge concerning their cognitive strategies as well as their knowledge about self and tasks. The questionnaire’s standardized answers make it simple to compile data.

MK score was defined as the sum of the items’ scores belonging to the MK in the GMQ. The maximum score was 40. Pre-service teachers who rated MK items “not true for me” or “hardly true for me” scored a maximum of 16 and were identified as pre-service teachers having low MK. On the other hand, pre-service teachers who rated MK items “rather true for me” or “very true for me” could obtain a minimum of 32 or higher and were identified as pre-service teachers with high MK.

Descriptive statistics were performed on the interaction types, the discussion scores, and message levels to determine any differences among the pre-service teachers whose level of MK are low, low-to-medium, medium-to-high, and high. Moreover, a multiple regression and correlational analysis (MRC) was conducted to establish the relative predictive importance of pre-service teachers’ MK scores in their discussion performance in WBAC when gender, domain of interest, ages, and prior success were controlled. As the previous studies reported that the gender (Howland & Moore, 2002), domain of interest (Hislop, 2000), age (Jiang & Ting, 1998) and prior success (Hew & Cheung, 2003) might have an effect on the pre-service teachers’ performance in WALE, these were included as confounding factors. Other contextual factors such as the level of web literacy, cognitive learning styles, etc., were not included in the study because of the small sample size. Increasing the independent variables decreases the power of the study.

Results

The number of messages posted in the forum discussions among the pre-service teachers and the instructor was 1,055. The mean value of the number of messages for each chapter was 3.39 per pre-service teacher. Pre-service teachers took a discussion score for each week out of 10, which was the mean of the scores received from the messages forwarded on that week. These scores were totaled at the end of the semester to obtain students’ total discussion scores. The mean of the total discussion scores was 69.21 out of 100. Standard deviation was 12.36. The range of the total discussion scores of the pre-service teachers was between 30 and 90. From this can be deduced that messages had moderate-to-high quality of participation in the asynchronous online discussions.

<table>
<thead>
<tr>
<th>Interaction types</th>
<th>Sample messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgment</td>
<td>“I agree with student X on some points. She says that presenting a question in a visual form makes it easier comparing the presentation of it in other forms.”</td>
</tr>
<tr>
<td>Question</td>
<td>“Maybe the teacher communicates with students well enough, but what will happen when the teacher does not present the course well? Students do not learn in this way.”</td>
</tr>
<tr>
<td>Compare</td>
<td>“I think expository teaching and discovery learning are different from each other in some aspects. Firstly, expository teaching is a cause for learning as it is understood by the meaning. Teachers use expository teaching method when they present a lesson. And I think expository teaching is an effective teaching method to get a good result for learning. We can say that the expository teaching method is like deductive reasoning. In both of them, the lesson goes from general to specific. Students are not so active; but they are not completely passive either…”</td>
</tr>
<tr>
<td>Contrast</td>
<td>“First of all, one of the major differences between the first and the second episode is that, while the first episode is inductive, the second one is deductive. The first one is inductive because students try to reach a general result from the specific patterns with the help of the teacher. Concretely, the teacher starts a discussion in which students give their ideas about the relations of the numbers in these operations considering whether they are odd, even or prime. Lastly, they end with the general rule that “an even number greater than 4 can be written as addition of”</td>
</tr>
</tbody>
</table>
two odd prime numbers.”

**Evaluation**

“I think most effective characteristic of this drama is asking the question in a different way. The teacher wants students get inside of the situation and find a solution. As in the example, the student tells the teacher, “I sent 20 more brown cows to drink water so there were equal numbers of cows and sheep.” Now we see that student considers himself as shepherd, thought and found a way to provide equality. But we can guess from this that each student can have different point of view for this question. And this view is exactly related to the understanding of the question…”

**Idea to example**

“…In order to be accountable for his/her work, there must be methods to assess individual work. The group’s performance is based on the individuals’ contributions. Each student receives a grade on his/her portion of the team project or product and the contribution of each individual is made known to the team…”

**Example to idea**

“I think questions with yes/no answers are ineffective. Because students have a 50% chance to answer it, so without thinking s/he makes up the answer quickly. For example, some teachers ask ‘We’re doing ______, aren’t we?’ This is a kind of meaningless question, as is ‘What is the answer? Is it x = 5?’ Ineffective questions are nothing but a waste of time; teachers should prepare their questions before the lecture and ask them at the right time.”

**Clarification**

“In cooperative learning, each learner is held accountable for his or her own learning and is motivated to increase the learning of others. This means that other members of the group are responsible for the slow learners. In this sense, heterogeneity at the achievement level is very important because it maximizes peer tutoring. With a higher achiever on each team, the coordination among the group members and the introduction of new materials become easier. Low achievers have opportunities to contribute to the group and seek clarification when needed…”

**Cause-and-effect**

“...In my opinion, questions are the tools that teachers use in order to open students’ minds, make them think, and increase their cognitive processes. Question should be clear and simple and cause students to think but not make them say “I do not know.” Because when it is difficult or complex, one student prefers the simple way, which is saying I do not know or I have no idea. Bloom also thinks in the same way. He says, ‘questions are our frequently used tools to explore our pupils, but we should not use them too frequently.’ Actually, it is a question-answer type class then, which is not a lesson but rather a contest.”

<table>
<thead>
<tr>
<th>MK Levels</th>
<th>MK Scores</th>
<th>n</th>
<th>A</th>
<th>Q</th>
<th>CM</th>
<th>CN</th>
<th>EV</th>
<th>D2E</th>
<th>E2I</th>
<th>CLE</th>
<th>CE</th>
<th>OT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>&lt; 16</td>
<td></td>
<td>4</td>
<td>30</td>
<td>31</td>
<td>6</td>
<td>4</td>
<td>28</td>
<td>0</td>
<td>16</td>
<td>7</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Low-to-medium</td>
<td>16–24</td>
<td>7</td>
<td>31</td>
<td>29</td>
<td>10</td>
<td>13</td>
<td>35</td>
<td>22</td>
<td>31</td>
<td>32</td>
<td>17</td>
<td>5</td>
<td>225</td>
</tr>
<tr>
<td>Medium-to-high</td>
<td>24–32</td>
<td>17</td>
<td>37</td>
<td>24</td>
<td>26</td>
<td>28</td>
<td>54</td>
<td>86</td>
<td>143</td>
<td>81</td>
<td>51</td>
<td>5</td>
<td>535</td>
</tr>
<tr>
<td>High</td>
<td>&gt; 32</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>11</td>
<td>18</td>
<td>11</td>
<td>38</td>
<td>35</td>
<td>28</td>
<td>16</td>
<td>0</td>
<td>165</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>102</td>
<td>88</td>
<td>53</td>
<td>63</td>
<td>128</td>
<td>146</td>
<td>225</td>
<td>148</td>
<td>90</td>
<td>12</td>
<td>1055</td>
<td></td>
</tr>
</tbody>
</table>

The distribution of the pre-service teachers’ messages according to the interaction types and their MK levels is presented in Table 4. The distribution of the pre-service teachers’ total discussion scores and MK levels for each chapter is illustrated in Table 5. Pre-service teachers having low MK had mainly low discussion scores and their messages contained mostly “acknowledgements,” “unsubstantiated judgment,” or “thoughtful query” type interactions. On the other hand, pre-service teachers having high or medium-to-high MK had mainly high discussion scores and produced mostly “example-to-idea,” “clarification and elaboration,” or “idea-to-example” type interactions.

Throughout the forum, responses were coded as initial posting (Level 1), response to the initial post (Level 2), reply to a response (Level 3), reply to a reply (Level 4) and so on. Of those 1055 messages, 494 messages were at level 1, 333 messages were at level 2, and 172 messages were at level 3. In total, 5% of the messages were at level four or above. This is indeed significant. However, “conversation” among participants did not have sufficient depth. CM
and CN feed the continuity of the conversation, but CM- and CN-type interaction is only 5% and 6% of the interactivity, respectively. For each MK level, Table 6 shows the percentage of message levels. Pre-service teachers having high MK mostly initiate interactions (Level 1) whereas pre-service teachers having low MK usually sent second-level messages.

**Table 5. Distribution of the number of pre-service teachers according to their discussion scores and MK levels**

<table>
<thead>
<tr>
<th>MK Levels</th>
<th>MK Scores</th>
<th>Discussion Scores (out of ten)</th>
<th>n</th>
<th># of pre-service teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>&lt; 16</td>
<td>&lt; 6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-to-medium</td>
<td>16–24</td>
<td>&lt; 6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium-to-high</td>
<td>24–32</td>
<td>&lt; 6</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>&gt; 32</td>
<td>&lt; 6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prior to conducting MRC, potential covariates — age, department, gender and GPA — were correlated with the total discussion scores in order to determine which of these should be considered as covariates. The results of these correlations and their significances were as follow: Correlations of the GPA, age, gender, and department with total discussion score were .509, −.326, .291 and −.333, respectively. The GPA had significant correlations with the total discussion score at the .01 level (2-tailed). However, gender, age, and department did not have significant correlations with the dependent variable at the .01 level (2-tailed). Hence, the GPA was determined as a covariate for the MRC.

Covariant was set to Block A, group membership was set to Block B, and interaction was set to Block C. Then MRC was performed to test the significance of R^2 change using the enter method for the dependent variable. The contribution of Block C is not significant for the dependent variable, (F [1, 27] = 0.306, p = .585). This result indicated that there was no significant interaction between the covariate and the group membership; therefore interaction (Block C) was dropped. For total discussion score, contribution of Block B is significant (F [2, 30] = 12.59, p = .000). MK score explained 21.4% of the variance in the total discussion scores when pre-service teachers’ GPA was controlled.

**Table 6. Percentages in message levels according to the MK levels**

<table>
<thead>
<tr>
<th>MK Levels</th>
<th>Message Levels</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Low</td>
<td>17</td>
<td>51</td>
</tr>
<tr>
<td>Low-to-medium</td>
<td>25</td>
<td>43</td>
</tr>
<tr>
<td>Medium-to-high</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>High</td>
<td>48</td>
<td>24</td>
</tr>
</tbody>
</table>

Conclusions, discussion, and implications

The relationship between discussion scores and MK level implies that MK has predictive value in the quality of the participation in the asynchronous online discussion. The statistical result of the SPSS reported R^2 change as .214 for total discussion score. Effect size for the total discussion score was 0.27, which matched a large effect size. In other words, this result has practical significance for similar populations. As a result, pre-service teachers’ MK was an important determiner of the variance in the quality of the participation, which refers to the degree of the engagement in a meaningful discussion.

Frequently used interaction types in the forum discussions were “example to idea,” “clarification and elaboration,” or “idea to example.” The high occurrence of these types is not surprising when the nature of the facilitative/guiding
questions is taken into account together with the pre-service teachers’ high MK scores. This reveals that pre-service teachers with high MK scores were more aware of the content of the messages and therefore made the connections between the ideas in the various messages and their own cognitive states. In other words, they were more context-aware than pre-service teachers who exhibited low MK. From this finding we can deduce that MK might have a facilitative role in constructing a context-aware ubiquitous learning environment. This study suggests that future researchers might deal with training of learners in MK so that learners can achieve context-aware ubiquitous learning.

Pre-service teachers rarely used compare and contrast messages. In fact, there was an insufficient amount of guiding and administrative elements in the messages, which affected the quality of the participation. Eventually, this resulted in a low level of interdependence and interrelations among pre-service teachers. In other words, message content did not show enough growth in discussion and the forum progressed on the basis of an accumulation of monologues. This constructivist online course did not produce effective enough collaborative learning environments, which require two or more individuals sharing understanding and providing mutual feedback. Only 5% of the messages were level four or above. It was expected that the pre-service teachers would have deepened the existing discussion by posting cause-and-effect or compare-or-contrast messages. However, pre-service teachers contemplated creating messages for drawing their experiences, referring course material, and introducing completely new information for discussion. Furthermore, they failed to evaluate themselves as cognitive agents in their connection with the task, because they usually preferred to discuss the topic as if they were outside of the case unless a prompt engaged them in the case or made them a part of the case. Future researchers might investigate the features of the messages that lead to coherent and deep discussions in the online discussion forum.

Another discussion is that pre-service teachers who rated low for MK items, such as “I check over my study to make sure I did it right” or “When I study, I put the important ideas into my own words,” made less effort to reach monitoring accuracy. Related to this, they also experienced problems navigating the course site and using help tools that could have given them important information to actively participate in the discussion. In fact, they were not aware of how much information about the tools was available on the course site and they did not make sufficient effort to develop knowledge of the task at hand. This finding also agrees with that of Ahitikari and Eronen (2004), who reported that web literacy is an important determiner for interactivity as well as participation in the activity on the course site. So, course site designers in WALE should encourage students to use search tools such as Google, Yahoo, and other in-course search tools and to follow useful links in the course site. They even should provide some key words depending on the context and new information or experiences to provide checkpoints for knowledge of self and task and to deepen the discussions. An important research implication is that instructors should include tools or activities that cause scaffolding of students’ thinking; prompts that cause judgment about “knowledge of self” (e.g., “Your message has been responded to three times. How can you defend or enrich your point of view?”) and stimulate interactive and rich messages and the effective use of course site tools. Secondly, instructors should include activities that force students to assess their learning needs, for example, “Which strategies do you plan to use for this case?” or “Do you have an idea about which of your friends might need help?” Forum discussions should thus involve instructive and guiding activities that enhance the use of MK strategies.

The study implies that instructors should encourage students to send messages explaining or clarifying concepts using examples and/or metaphors. Throughout the transcripts of the forum messages, the pre-service teachers posed 225 example-to-idea type messages, which contained 213 examples and 12 metaphors. Forty-six example-to-idea messages (out of 213 example-containing messages) were followed by the cause-and-effect messages. Forty-one percent of them were produced by the pre-service teachers having low or low-to-medium MK. Therefore, it can be inferred that examples or metaphors positively influence participation of the students having low MK by awakening their knowledge of self and task. The benefits of using examples or metaphors might be two-fold: (1) followed messages, possibly, might contain high levels of interactions, which are required or wanted in any online discussions and (2) metaphors or examples might push students to control and check their knowledge structure in their mind related to concepts under discussion. Besides, these benefits might explain why pre-service teachers with high or moderate-to-high MK mostly initiated interactions whereas pre-service teachers with low or low-to-moderate MK scores mostly sent the second-level messages.

Metaphors or examples stimulate awareness of knowledge of task and self, explaining why pre-service teachers with low MK scores engage more in meaningful discussions. Eventually, there are two important consequences. One is that active awareness of MK strategies encourages students to exemplify what they think or to offer metaphors about
what they understand. Consistent with Kanuka and Anderson (1998), the second consequence is that creating meaningful discourse among the students requires students to be able to demonstrate critical thinking skills by relating course content to prior knowledge and experience, and making inferences. Consequently, students who are in an instructional medium that stimulate MK skills can effectively participate in the discussion because they previously predict their learning needs by using their cognitive domain.

The implication of this study for educators involves the importance of MK on effective participation to the discussions. Educators should begin to focus on the MK of the learners and foster the development and use of it when necessary. From this perspective, future research should investigate the relationship between the other components of the metacognition and coherent and deep online forum discussion by considering the learners’ needs/requirements with regard to online teaching/learning in specific contexts. In this way, asynchronous online learning environments’ demands for autonomous learners might be met and a remedy for difficulties stemmed from the discontinuity in WALE might be found.

References


Applying Web-Enabled Problem-Based Learning and Self-Regulated Learning to Add Value to Computing Education in Taiwan’s Vocational Schools

Tsang-Hsiung Lee
Department of Information Management, National Chengchi University, Taiwan // Tel: +886-2-2939-3091 ext. 81243 // Fax: +886-2-2234-8444 // design2learn@gmail.com

Pei-Di Shen
Department of Information Management, Ming Chuan University, Taiwan // Tel: +886-2-2882-4564 ext. 2212 // Fax: +886-2-2881-0521 // pdshen@mcu.edu.tw

Chia-Wen Tsai
Department of Information Management, Yuanpei University, Taiwan // Tel: +886-2-8631-7134 // Fax: +886-2-2880-9764 // cwtsai@mail.ypu.edu.tw

ABSTRACT
This article describes the design and delivery of a compulsory course in packaged software at vocational schools in Taiwan. A course website was devised and deployed to supplement learning activities in the traditional classroom. A series of quasi-experiments was conducted with innovative instructional designs, that is, web-enabled problem-based learning (PBL), self-regulated learning (SRL), and their combinations. The impacts of these web-enabled pedagogies on students, instructors, and course design were evaluated. The effects of web-enabled pedagogies were mostly positive, thus reinforcing the instructor’s confidence for further application to the rest of his courses. The authors further discuss the implications for schools, scholars, and teachers who plan to implement, or are already engaged in, e-learning.

Keywords
Web-enabled PBL, Web-enabled SRL, E-learning, Computing education

Introduction
It has become tougher and much more competitive in recent years for vocational schools in Taiwan to attract sufficiently high student enrollment. This is mainly the result of a continually decreasing birthrate and, at the same time, a rapidly increasing number of new schools. While students graduating with vocational degrees represent a significant proportion of the higher education population, they, on average, receive a much lower public investment compared with their counterparts in the universities. A typical student in a vocational school tends to have a lower level of academic achievement and spends more time on out-of-class jobs. These students also do not get adequately involved in their schoolwork and tend to care less about their grades. The above conditions challenge almost every teacher in this sector who tries to help develop students to be more competently skilled and competitive in the labor market.

No one doubts the guiding principles of practical applications in vocational education in Taiwan (Tai, Chen, & Lai, 2003). However, most of the teaching and learning efforts in this area have been devoted to helping students pass written tests in order to receive awards or official certificates. Schools, feeling the increasing pressure of competition, often emphasize the proportion of students awarded such certificates before they graduate instead of the quality of learning. This materialist approach results in students’ attention being less on practicing skill applications and more on preparing for tests through memorization. Consequently, students who have passed an examination may still be unable to apply what they have learned in school, and worse, they may lack the motivation to learn more in the future.

Computing courses traditionally emphasize memorization by applying short, disjointed, lack-of-context examples. This results in a gap between what is learned in school and what is required in the workplace (Wu, 2000). In this regard, the computing education in vocational schools in Taiwan can hardly be deemed as effective. In order to increase students’ learning motivation and to develop practical skills, problem-based learning (PBL) is considered to
be one of the most appropriate solutions. PBL uses real-world, simulated, contextualized problems of practice to motivate, focus, and initiate content learning and skill development (Boud & Feletti, 1991; Bruer, 1993; Williams, 1993). We believe that PBL would help less academically inclined students to develop practical computing skills.

Web-enabled instruction seems to be an ideal learning environment because students can access an almost unlimited amount of information and apply it in multiple ways (Kauffman, 2004). However, implementing e-learning for lower academically achieving students inherently has high risks. For instance, Internet addiction is quite common among these students. When students enter the traditional classroom, they are used to logging on to MSN messenger and checking their e-mail first. Many students like to chat with each other frequently via MSN messenger even when they are in the same classroom. They might browse shopping websites or play online games while the teacher is lecturing in the class. It is even more difficult for students to concentrate on online learning because of this addiction to the Internet and the lack of on-the-spot teacher monitoring. To respond to this challenge, the authors turned to an approach that can help students better regulate their learning.

Success in online courses often depends on students’ abilities to successfully direct their own learning efforts (Cennamo, Ross, & Rogers, 2002). It is very critical to develop students’ regulation of learning before providing online courses to them. This particularly applies to low academic achievers. Students’ motivation may benefit from web-enabled instruction with self-regulated learning (SRL) strategies. Students in the online environment who are equipped with SRL competence become more responsible for their learning and more intrinsically orientated (Chang, 2005). Successful students in an online course generally used self-regulated learning strategies, and the effect of the self-regulation variables on students’ success was statistically significant (Yukselturk & Bulut, 2007). Consequently, SRL was applied in this study to help vocational school students concentrate on their learning, to leave time for learning after their out-of-class jobs, and furthermore, to take responsibility for their learning.

This study reports a learning journey of transitions experienced by an instructor. The instructor has taught a course called Packaged Software and Applications for many years, and held a strong belief that traditional teaching methods are still the best. This belief was built largely from his past success in helping students pass licensure exams and thus receive certificates. However, a closer look at students’ performance reveals that receiving certificates does not equate to quality learning. When the above web-enabled pedagogies became plausible alternatives, the instructor’s advisors suggested that he try them in his classes. Instead of directly translating the teaching materials into electronic form, the instructor redesigned the course and conducted a series of quasi-experiments to examine the effects of web-enabled PBL, SRL, and their combinations.

There are few studies that have discussed effective online instructional designs for low academic achievers. In this area, the restructuring and translation of traditional computer software courses into course websites has seldom been documented. This study is the first trial for this instructor to implement innovative instructional designs, including PBL, SRL, and e-learning, in his classes. A steep learning curve is inevitable. The adaptation of innovative instructional designs and new technologies challenges not only the teacher but also the students. In this study, the authors provide the valuable experience of deploying web-enabled pedagogy by this teacher, who is an expert in traditional teaching methods but a novice in innovative teaching methods, to foster student learning through the Internet. Looking at the results, we believe that experiences learned through this implementation of web-enabled pedagogies are worthy of consideration by other teachers who plan to implement, or are already engaged in, e-learning.

**Research approach and study settings**

In this study, the main goal was to improve students’ learning, explore the web-enabled learning effects, refine the online course, and reinforce the teacher’s professional development. Therefore, mixed methods were applied. Specifically, the case study method was adopted to describe how students and the teacher adapted to the online course, and to detect deficiencies in design ideas and associated implementations for possible remedies in future trials. The authors intervened in students’ learning via different instructional designs and deployed quasi-experiments to explore the effects of the intervention.
Course design

A required course in a vocational school, Packaged Software and Applications, focuses on the development of students’ computing skills in applying packaged software. The course under study is a semester-long, two-credit-hour course, targeted at first-year college students with majors in different areas. A credit hour in Taiwan is defined as 16 hours of instruction (including exams) over the period of one semester, so in a two-credit-hour course, students receive 32 hours of instruction. Upon successful completion of the course, as measured by exams, papers, and project work, a student will be awarded a grade valued at two credits. Students solve a series of tasks by applying Microsoft Office (including Word, Excel, and PowerPoint). There were two classes in this study. One was a PBL class, and the other was a non-PBL class. In addition, each class had two groups divided according to whether the students were involved in SRL or not. Therefore, there were four groups (conditions) in this study (see Figure 1).

![Figure 1. The expected effects of different instructional designs](image)

Learning activities

The course was divided into three sequential modules: the Word module, the Excel module, and the PowerPoint module. A skill test was held after the completion of each module. The first test was held during the midterm examinations, the 8th week, the second test was held in the 13th week, and the final one in the 16th week. A detailed schedule of the experiment is depicted in Figure 2.

![Figure 2. Schedules of the three modules and skill tests](image)
In the beginning of the course, students were encouraged to adapt and learn via a course website. Teaching in this period took place in the traditional classroom. The teacher first recorded every session of his lectures and then translated these lectures into HTML files with flash, video, and voice. These HTML files were then loaded onto the course website. Students could preview and review the course sessions via this course website. After three weeks, most of the coursework was moved into the website. The instructor helped students adapt to learning on the net and lessen the feelings of isolation. Within these three weeks, the instructor adjusted students’ learning gradually and smoothly. On this course website, students got support not only from the instructor, but also from their peers.

**Design of the course website**

This was the first attempt by this instructor to adopt innovative teaching technologies, including a course website and audio-recording software. Before that semester started, the instructor searched for an appropriate online teaching platform. The content management system, XOOPS (eXtensible Object Oriented Portal System), was adopted as the course website. XOOPS is an open source framework and provides functions of user management, anonymous discussion forum, file downloading, file uploading, website links, and group management (see official website at: http://xoops.org).

The course website mainly consisted of four sections: course information, course content, course discussion, and student system. “Course information” provided course description, syllabus, assignments, grading, and course-related information; “course content” included the audio files and the examples for students’ exercises. Students could download the files and listen to the recording via the website to review or complete exercises repeatedly. The instructor could ask questions in the course discussion board in order to promote discussion and interaction between students and himself. In addition, students in SRL groups were asked to regularly keep learning journals. These journals were located in the student system.

**Design of quasi-experiments**

The experimental design was a 2 (PBL vs. non-PBL) × 2 (SRL vs. non-SRL) factorial pretest/post-test design. The participants were randomly assigned to one of the four experimental conditions; each group contained about 30 participants. However, as mentioned above, vocational students in Taiwan tend to have lower level of academic achievement and involvement in learning, and care less about their grades. They give up on their learning easily. Twelve students withdrew from this course during the instructional process. At the end, 102 students remained in the classes.

Students in the four groups solved the same tasks but in different learning conditions. Students in the PBL and SRL group (C1, n = 28) were required to prepare and read the textbook, papers, and related documents before classes. Furthermore, they were taught to solve a problem or build a business file in an authentic situation. The PBL and non-SRL group (C2, n = 25) focused on the teaching effectiveness in an authentic situation without the teacher’s help focusing on forming self-regulated learning habits in these students. They just participated in this course by listening to the teacher; however, they were not asked to preview or review the content of classes. The non-PBL and SRL group (C3, n = 24) was taught in a traditional teaching style, but students were required to prepare the course material before classes. Groups C1, C2, and C3 were experimental groups, while the non-PBL and non-SRL group (C4, n = 25) was the control group. Students in group C4 were treated and taught in the most traditional way without any simulated problems or requirement for self-regulated learning. That is, teaching this group still emphasized traditional memorization by applying short, disjointed, lack-of-context examples, although the content was delivered via the Internet.

**PBL Treatment**

The teacher created interesting, challenging, and authentic problem situations. In the first Word module, students were required to apply for a job as marketing assistant for an online-game company. They were required to design and then build autobiographies and resumes by applying computing skills they had just learned. In the Excel module, students played roles as if they were employed by this same software company, and a marketing manager asked them to compare expenses resulting from different distribution channels. They had to survey data and then complete a
worksheet with graphs to contrast the differences between channels. Additionally, they had to come up with a recommendation regarding the best combination of channels. In the last module, they were promoted to marketing managers. They were asked to develop a business proposal for a new online game. They had to present this proposal with visual aids to convince the managing director to enter the market. Therefore, a persuasive PowerPoint file was built into this phase. In summary, authentic and continuous problems were applied in the PBL class throughout the semester.

Teaching in each of the three modules proceeded similarly, as follows: First, the teacher would demonstrate how to approach the situation and solve the problem through web-enabled multimedia. In addition to teaching computing skills, the teacher also discussed similar situations and related applications. In the latter, teacher would guide students in constructing their own models for problem-solving.

**SRL Treatment**

There was a SRL group for each class. Students in SRL groups received an after-class course teaching them SRL strategies. The two SRL groups from the PBL class and non-PBL class gathered in a classroom for a two-hour lecture discussing how to manage study time and regulate their learning. The content of this SRL course was composed of the four processes addressed by Zimmerman, Bonner, & Kovach (1996), including self-evaluation and monitoring, goal-setting and strategy planning, strategy implementation and monitoring, and monitoring of the outcome of strategy. Students were taught how to implement these four processes to become more self-regulated learners.

In addition to the two-hour lecture, students in the SRL groups were required to regularly prepare and read the textbook before classes, and to review or practice the computing skills they had learned after class. They were also required to record their learning behavior every week. The data was recorded on the course website instead of in their notebooks in order to prevent falsification. The teacher scanned students’ records. The treatment among the four groups is illustrated and compared in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>Teaching Activities</th>
<th>Learning Activities</th>
</tr>
</thead>
</table>
| C1    | The teacher…
|       | • demonstrated how to solve authentic problems and discussed potential applications.
|       | • taught SRL skills and urged students to study regularly. | The students…
|       | • took on authentic tasks and learned by problem solving.
|       | • practiced SRL and recorded learning behaviors every week. |
| C2    | The teaching activities were the same as C1 but without SRL lectures. | The students experienced authentic situations and solved the problems without extra requirements of SRL. |
| C3    | The teacher…
|       | • converted his traditional way of teaching into an online format without any other modification.
|       | • taught SRL skills and urged students to study regularly. | The students…
|       | • received the traditional computing course through the Internet. | • practiced SRL and recorded learning behaviors every week. |
| C4    | The teaching activities were the same as C3 but without SRL lectures. | The students experienced the traditional style of teaching and did not deal with the extra requirements of SRL, although teaching was conducted via the Internet. |

**Data collection and measurement approach**

Through a pretest/post-test, quasi-experimental design, the researchers empirically assessed the differences in the four groups at the beginning and end of the study. Therefore, the authors measured students’ computing skills and
involvement before the course started. In Taiwan, almost every student learns Microsoft Word before learning any other packaged software. That is, a student without the computing skills of Word may be considered extremely limited in skills for other application software. The experiment conducted in this study was to measure students’ improvement in Microsoft Office (including Word, Excel, and PowerPoint) with Word as the first module in this course. Therefore, it is believed that Word could be a predictor of knowledge or skills in Microsoft Office. In this regard, the instructor first measured students’ computing skills using Word as a baseline at the beginning of the course. Students completed three Word documents as a pretest, and the score showed a uniformly low skill level. The difference among students’ computing skills on the pretest among the four groups was not statistically significant. This confirmed that all participants in the four groups had low levels of knowledge and skills involving packaged software. Thus, the researchers ruled out initial differences as a plausible alternative explanation for the differences detected after treatments (Gribbons & Herman, 1997).

Moreover, Zaichowsky’s (1985) Personal Involvement Inventory (PII) was used in this study to measure students’ involvement at the beginning of the online course. The definition of involvement used in constructing the PII had much in common with motivational theory (Schmidt & Frieze, 1997), and measured three constructs: interests, needs, and values. In order to avoid uneven distribution of students among the groups in regard to level of involvement and motivation in this new course that combined innovative instructional designs and technologies, the authors tested students’ involvement before the course started. The differences among the four groups in student involvement in the course at this beginning stage were not statistically significant. Therefore, it is considered that the students had equal levels of skill in Microsoft Office and involvement when they began this course. In addition, none of them had any experience in taking a web-enabled course. The authors then evenly and randomly divided the students into the four experimental groups.

At the end of each module, a skill test was given to students. Before testing, students were assigned to random seats. All students were tested at the same time. The questions in the test related to the content and examples in the course. Every test consisted of 5 to 7 questions. The teacher graded and recorded the results immediately after each test. A surrogate grade representing computing skills was averaged from the scores of these three tests. Finally, the enhancement of computing skills in a module was the result of one’s surrogate grade minus his pretest grade. The authors tested the differences in the enhancement of computing skills under different conditions.

In addition to the quantitative data, qualitative data were also collected. During the semester, students in the two SRL groups were required to record their reading, previews, and reviews as evidence of their self-regulated learning activities on the teaching website. After the course was completed, research assistants collected the data into tables. At the end of the course, the instructor in this study interviewed eight students from the four groups. Two students were chosen randomly from each of the four groups. The interviews were recorded, transcribed, and analyzed.

In this study, the same instructor was responsible for teaching and overseeing the four groups. The impact of the implementation of web-enabled pedagogy on the instructor was investigated through his own journal entries. The same instructor also tracked critical incidents throughout the semester.

Outcomes

Impact on students

As the quantitative data show in Table 2, the enhancement of students’ computing skills in terms of their average grades on three modules (Word, Excel, and PowerPoint) in the PBL class (67.48) was significantly higher than that in the non-PBL class (57.00) \( (p = 0.000) \). That is, in a web-enabled learning environment, the effects of problem-based learning on enhancing students’ computing skills may be positive, and higher than of those without PBL.

Moreover, the results from Table 2 show that the enhancement of computing skills in the SRL group (66.39) was significantly higher than that of the non-SRL group (58.34) \( (p = 0.001) \). This supports the claim that the effects of web-enabled SRL on students’ computing skills may be positive, and that students with SRL may perform better than those without SRL.
Table 2. Independent sample t-test: Improvement of grades

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>S. D.</th>
<th>F</th>
<th>t-value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Averaged scores of three skill tests</td>
<td>PBL</td>
<td>53</td>
<td>67.48</td>
<td>9.248</td>
<td>7.913</td>
<td>4.760</td>
<td>100</td>
<td>.000***</td>
</tr>
<tr>
<td>minus the score of pretest</td>
<td>non-PBL</td>
<td>49</td>
<td>57.00</td>
<td>12.821</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SRL</td>
<td>52</td>
<td>66.39</td>
<td>10.978</td>
<td>1.647</td>
<td>3.501</td>
<td>100</td>
<td>.001**</td>
</tr>
<tr>
<td></td>
<td>non-SRL</td>
<td>50</td>
<td>58.34</td>
<td>12.233</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Finally, data from Table 3 show that the combination of PBL and SRL resulted in that group having the highest grades among the four groups in this research. The enhancement in computing skills of students in C1 was higher than those of C3 and C4 in a statistically significant manner (p < 0.05), and also higher than C2, though insignificantly. Thus, it may be concluded that students’ computing skills were improved by simultaneously applying web-enabled problem-based learning and self-regulated learning.

Table 3. One-way ANOVA: Average of the improvement of grades

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(I) Group</th>
<th>(J) Group</th>
<th>Mean Difference (I – J)</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same as appeared in Table 2</td>
<td>1</td>
<td>2</td>
<td>6.735</td>
<td>2.891</td>
<td>.151</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>9.238(*)</td>
<td>2.923</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>-6.735</td>
<td>2.891</td>
<td>.151</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2.503</td>
<td>3.003</td>
<td>.874</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>-11.160(*)</td>
<td>2.972</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>-2.503</td>
<td>3.003</td>
<td>.874</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>-17.895(*)</td>
<td>2.891</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>-11.160(*)</td>
<td>2.972</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>-8.657(*)</td>
<td>3.003</td>
<td>.046</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level with Scheffe test.

As revealed in the interview data, most students in the PBL class had received a spoon-fed computing education in their senior high schools. Despite having taken computing courses for three years, they still could not apply skills and knowledge learned to real-life problems. They did not know when and how to apply what they had learned in previous classes. A student who had successfully earned related certificates during her senior high school years made the following comments:

I had difficulties in “applying it” rather than “learning it.” Many times, I had no idea how to solve the problems, nor how to find ways in dealing with them, although I had taken computing courses and received the related certificates for many years. The traditional teaching materials and examples did not contribute to problem-solving when we faced the problems.

After the intervention of PBL, students agreed that this series of authentic problem-solving exercises had raised their interest and helped them understand the situations and how to apply their acquired skills. One student shared his thoughts regarding the new experience gained in learning:

This course provided special learning experiences that differed from those of the past computing courses. The simulated problems and the problem-solving processes helped us realize and reflect on how, where, and when to apply what we learned. For example, being a student in the department of business administration, I know how to compare different expenses by using a table and illustrate them with graphs after this course…The interesting examples associated with real situations also helped me concentrate on learning in the online course.
In contrast with students’ welcome of the implementation of PBL, students experienced difficulties and persistently struggled to self-regulate their learning. In the interviews, most students in the SRL groups admitted that they were still not used to the practices of SRL, such as preview before class and review after class. Very few students studied and practiced regularly. The students were not willing to spend extra time on their coursework after school, and resisted taking responsibility for their own learning. For instance, many students in the SRL groups did not prepare, practice, and record their learning within the deadline. Worse, three students in the C1 group even dropped out of this group because they did not want to spend more time on this course after school. One of the participants in the SRL groups described the processes of regulating his learning:

Although the teacher asked us to learn and record our learning in the regular time, we did not complete it in time…We have received a spoon-fed teaching method since we were elementary students. It is difficult for us to change our learning styles, that is, preview, review, and practice regularly, within a short time. Besides, it is a big challenge for those students with out-of-class jobs to be more involved in their learning. For these students, few learning tasks and records were done within the pre-specified time span.

However, the teacher still encouraged the students in SRL groups and insisted that students should persist in regulating their learning. At the end of the course, their achievement suggested that this effort was valuable. Interestingly, students in the four groups expressed that they would like to take more online courses for the rest of their school years. They felt comfortable and unrestrained in this online course. E-learning seemed to be more popular for those students who did not get adequately involved in their schoolwork and cared less about their grades mainly because of the conveniences and advantages of web-enabled learning. This unexpected finding impressed the instructor. Although the results of implementing SRL in this study were mixed, we may conclude that e-learning is not only possible but also effective for less academically motivated students, if appropriate pedagogies are well designed and implemented.

**Impact on instructor**

As mentioned before, the focal instructor had previous success in helping students pass official exams and earn related certificates. Not sure about whether this success might be sustained, the focal instructor worried a lot in the beginning of his attempts to conduct these experiments about the indeterminate outcomes and that all his efforts might be wasted. In the phase of teaching Word, the instructor recorded the following comment in his teaching journal:

Students seem very stand-offish and unwilling to adapt to the new technologies and innovative instructional designs. They look unhappy when they are assigned tasks and asked to solve the problems. I think that they expect me to spoon-feed them in the way that they are used to and that I am also skilled in. However, I prepared for this online course, and designed the teaching website and course for several months. Will my efforts contribute to students’ improvement of computing skills, or just be wasted?

However, an unanticipated learning event occurred in his class and thus removed some of his doubts. In the third week, the instructor left for an unexpected meeting during class. Before the instructor left the classroom, the students in the PBL class were told to practice and design a show bill in groups. After the teacher left the class, students started to discuss what and how to design by applying what they had learned. When the instructor came back, the show bills presented by the students were very exquisite and attractive. This instance encouraged the instructor greatly, and the instructor was convinced that the effects of web-enabled PBL were very likely effective, even for low academic achievers.

As for SRL, the instructor experienced many hurdles to lead and guide students in their transition to becoming self-regulated learners. Section 3.1 described students’ unwillingness to change during the implementation of SRL. On numerous occasions, the instructor almost gave up on the intervention of SRL. The instructor recorded frustration in his journal:

Each time when I hear students’ moans and complaints about the extra homework and assignments, I really want to give up SRL intervention in the experiment. They seem to take these course requirements as merely burdens. Does the intervention of SRL harass both students and me? It seems that only I, the instructor, am struggling to improve students’ practical computing skills to
strengthen their competitiveness in the job market. Students’ resistance and carelessness have almost dried up my patience and enthusiasm.

Nevertheless, with encouragements from the instructor’s advisers, this instructor continued on the intervention. Eventually, based on quantitative data, students’ enhancement of computing skills in SRL groups was significantly higher than that of in the non-SRL groups ($p = 0.001$). This suggested that the effects of SRL among lower academic achievers could be significantly improved even by a limited amount of intervention such as a two-hour lecture regarding SRL at the beginning of the teaching program and later by students’ monitoring of their own learning. Thus, the authors conclude that teachers should maintain their confidence, even when the inevitable resistance is daunting, and encourage students to persist in SRL.

**Impact on course design**

Findings of the quasi-experiments showed that the enhancement of computing skills of students in C1 was significantly higher than that of C3 and C4 and also higher than that of C2, though insignificantly (see Table 3). For those teachers who wish to stick to traditional methods of teaching, directly translating their teaching materials into electronic form may not be a fruitful approach. Students in the control group (C4) received the poorest grades among the four groups, and the differences in grades among them were significant ($p < 0.05$). Students who have a lower level of academic achievement may suffer from less effective learning if they receive traditional instructional designs and materials online without reconsideration and redesign. Inappropriate instruction through the Internet that lacks teacher’s on-the-spot monitoring cannot provide positive effects. Moreover, it may even dampen the original effects achieved with the same content in a traditional classroom. In this regard, it is suggested that teachers should redesign their courses and then adopt appropriate instructional methods and technologies to fully exploit the benefits of deploying web-enabled learning environments.

It was found that the courses in computing skills were appropriate to be instructed via Internet. However, the innovation in teaching and associated experiments in this study was far from perfect. Many problems occurred during the one-semester intervention. For example, the design of the learning environment left much room for further improvement to attract students’ attention. The next version of PBL design could be more structured in procedures and more adequate in text writing. Teachers should invest significant efforts in this design and start to prepare the design well in advance of its first implementation. In addition, the instructor should frequently check students’ learning records and provide timely feedback and support for the SRL groups in the initial phase to assist students in regulating their learning.

**Discussion**

**Effects of web-enabled pedagogies**

It is important to know what kinds of learning activities may best engage students in learning through the Internet. Talay-Ongan (2003) advised that teachers should shift or adapt traditional lectures to web-enabled-PBL to achieve constructive alignment in online teaching. In the present study, web-enabled PBL was found to play a relatively positive role in enhancing students’ skills of deploying application software in contrast to those students instructed without PBL. It was demonstrated that PBL via e-learning facilitated development of students’ skills of applying computer software in general, and students of vocational schools in Taiwan in particular (see Table 2). The findings in this research were also similar to those in Chanlin & Chan’s (2004) study, which revealed that students in the PBL treatment group performed better than those from the control group in a web-based approach.

The importance of self-regulation for effective learning in Internet-supported instructional environments has been emphasized heavily in the literature (see Winnips, 2000). Providing students with opportunities to integrate their knowledge through web-enabled instruction may not be effective if they lack adequate skills needed to regulate their learning. Strategies that prepare students for the rigors of learning at a distance, which may increase the probability of retention and success, must be put into practice (Chang, 2005). In this regard, this study also found that the effects of SRL on enhancing students’ computing skills are positive compared to students’ skill development without SRL intervention (see Table 2).
Though the SRL effects were relatively positive, however, one should not be too optimistic about this finding. Challenges remain ahead. Vocational students in Taiwan are immersed in a culture of spoon-fed teaching and learning from the time they are elementary students. Thus, the requirement of SRL for low achievers should not be too rigorous, as they cannot get used to taking responsibility for their learning in a very short time. Excessive and overly strict requirements, without adequate communication, before applying SRL, may lead to great resistance and antipathy to the treatment. To resolve the dilemma of assisting students in developing SRL skills and preventing their antipathy in the new experiences in e-learning, a teacher should adopt those SRL strategies that are more instrumental and acceptable to the students, low achieving students in particular.

With respect to the effects of combination of instructional methods, the outcomes showed some support. As shown in Table 3, the results revealed that the effects of a combination of PBL and SRL intervention on enhancing students’ skills of deploying application software were positive and higher than for those who did not receive PBL or/and SRL, although the difference between C1 and C2 was not statistically significant. This result is consistent with other results in the literature. For example, Paris & Paris (2001) revealed that PBL facilitated SRL because it placed the responsibility on the students to discover information, to coordinate actions and people, to monitor understanding, and to reach goals. In Perels, Gürtler, & Schmitz’s study (2005) of mathematical problem solving, it was found that combined training in self-regulatory and problem-solving strategies was effective for enhancing self-regulatory competencies in solving problems. Moreover, Kramarski, & Gutman (2006) revealed that SRL students significantly outperformed the non-SRL students in solving problems in the procedural and transferal tasks regarding mathematical explanations in web-based learning environment. Taking the above evidence together, this study suggested that teachers should consider, for the good of student learning, applying PBL and SRL simultaneously to their courses in the context of e-learning, rather than deploying them singly.

**Unexpected effects of learning to change both teaching and learning**

Many instructors earnestly explore innovative approaches to employ technologies to enhance the quality of teaching and learning in higher education (Reeves, Herrington & Oliver, 2005). Some scholars conduct design research to meet these new challenges. Two innovative instructional methods and associated technologies were employed in this design research to help vocational students improve their computing skills. This journey of designing web-appropriate materials, climbing a stiff learning curve, and overcoming much resistance might pose challenges to both instructors and students before the achievement of promising, or even acceptable, results. The experienced and technologically inclined teacher who actually taught the computer courses in this study had entry barriers to adopting new technologies, and the barriers were even higher while transforming his fundamental teaching methods. The instructor experienced a difficult transition from being a traditional classroom teacher, who was confident about helping students pass exams and receive official certificates, to becoming an effective online teacher. This indicates that some other teachers, particularly those with limited technological exposure, may find this process similarly daunting.

The strength of the Internet is in delivering information directly to individuals; however, this may also be one of its greatest dangers. The applications of e-learning allow students to work on their assignments whenever and wherever they want (Schwieren, Vossen, & Westerkamp, 2006). Nevertheless, students retreating to the isolation of their computers may avoid school activities and course involvement, and instead be content with self-gratifying Internet entertainment (Treu & Belote, 1997). Studies indicate that vocational students were more Internet-addicted than general students (Yang & Tung, 2007). Therefore, it is a big challenge for teachers to help vocational students to be involved in an online course in an environment that is full of Internet allure with millions of shopping websites and free online games, and even MSN Messenger. It is difficult for students to concentrate on and be involved in online courses because of this addiction to the Internet and lack of on-the-spot teacher monitoring. In this regard, teachers should redesign their courses to attract students’ interest and help them to be more involved in the online courses. This study suggests that web-based pedagogies such as web-enabled PBL and SRL can improve students’ concentration on study and further contribute to student learning. Teachers and researchers can adapt the web-enabled PBL and SRL suggested in this study to their courses, or adopt other innovative instructional methods to elaborate on the benefits of e-learning.

With the students’ struggle and teacher’s persistence, the web-enabled pedagogies applied in this study not only contributed to students’ improved skills in application software, but also to this teacher’s professional development.
Further, after the first attempt of implementing web-enabled pedagogies, the preliminary results raised the instructor’s confidence and enthusiasm to extend the scope and depth of the experiment in the near future.

**Refinements for web-enabled pedagogies**

The instructor experienced many hurdles in leading and guiding students in the transition to becoming self-regulated learners. On numerous occasions, the instructor almost gave up the intervention of SRL because of students’ resistance. Some shortcomings that obstructed SRL in this attempt should be addressed. Firstly, when initiating web-enabled SRL, teachers should enforce the study requirement that all the students study or practice regularly and record their learning every week on the website to prevent their lagging behind in exercising this new strategy. Secondly, it was also very important for the teacher to monitor students’ learning via their learning record and to give timely feedback to increase students’ awareness about appropriate learning behaviors. Thirdly, the requirements of SRL for low achievers should not be too rigorous, as they could not get used to becoming responsible for their own learning within such a short period. Excessive and overly strict requirements of SRL without adequate communication beforehand may lead to greater resistance. Finally, lecture content about knowledge, action strategies, and exemplar cases of SRL should be strengthened to increase students’ willingness and thus confidence. As time goes by, they may have a better chance of becoming self-regulated learners.

With regard to the treatment of PBL, it is not always easy to apply the principle of jointed and related examples throughout the whole semester. However, the next version of PBL design shall be more structured in procedures and provide adequate scaffolding in text. Teachers should invest significant efforts in this design and start to prepare the design far in advance of its implementation.

**Limitations and opportunities for further research**

Though the quantitative and qualitative data both show positive effects of improving students’ computing skills, there were some limitations in drawing firm conclusions due to threats to the validity of the quasi-experimental design. A major problem with this design might be that the four groups were not be necessarily the same before any treatment or instruction took place, and might differ in important ways that influenced their performances. In this regard, the researchers empirically assessed the differences in students’ computing skills and involvement in this course among the four groups in the beginning of the study. The differences in students’ computing skills and involvement in the pretests were not statistically significant. The researchers could thus rule out initial differences as alternative explanations for the differences detected (Gribbons & Herman, 1997).

Some problems result from students in the comparison group being incidentally exposed to the treatment condition, having more enthusiastic teaching, being more motivated than students in the other group, etc. (Gribbons & Herman, 1997), which might influence the effects of online learning. The enthusiastic and involved teacher may engage more and pay more attention to the experimental groups, which may result in biases and exaggerate the effects of the adopted teaching methods. One should be aware of these contextual factors that may threaten the validity of claims made by this study.

Some other factors might also potentially influence the effects of students’ online learning. A student with readiness for self-directed learning may appropriately adapt himself to a web-enabled learning environment, resulting in better learning performance. For example, it was observed that several students with good grades in the traditional classroom recorded their learning regularly, and performed better than those without self-regulation. These students were relatively more self-directed in the traditional classroom instruction and also had better learning effects in the online environment. Moreover, students’ preference for computer courses might also lead to better learning effects. For instance, some students with preference for learning software showed conspicuous motivation in this online course. Students who preferred computer courses might adapt better to technology-based instruction and thus perform better. The instructor observed that some of his students still logged on the course website to review the content even after the semester had ended, and asked for other online courses in the coming semester. Altogether, factors such as students in the comparison groups with more motivation, more enthusiastic teaching, students’ past grades, preference for computer courses, and readiness for SRL, might influence the effects of online learning. To
complement the findings of this study, we advise further studies to explore the relationship between these contextual factors and students’ online learning.

Conclusion

Teachers face tremendous challenges in implementing e-learning in the environment where students are surrounded with free online games and shopping websites. So it is not immediately clear how to focus students’ attention, improve their learning, and help them be more involved in a web-enabled course without the teacher’s on-the-spot monitoring. In this study, we deployed two instructional designs, PBL and SRL, to enhance students’ learning through the Internet. Based on the results, experiences, and insights gained from this implementation, we believe that this research contributed in three different ways. Firstly, this study specified how teachers can engage students in improving learning under authentic conditions. At the same time, the teachers helped students to regulate their learning by applying PBL and SRL instructional methods in a web-enabled learning environment. Secondly, this study followed a traditional teacher who had success in helping students pass certification exams along his journey of adapting to the innovative instructional designs that he was adopting in his course for the first time. The results reinforced this teacher’s confidence in conducting online courses, particularly for students with lower academic achievement. Thirdly, this study took note of the shortcomings and insufficiencies that were found in the first implementation of the web-enabled pedagogies in this context and suggested refinements for the next experimental round.

This study may provide valuable insights and shed light on new and effective practices for schools (particularly vocational schools), scholars, and teachers who are planning to implement or are presently engaged in implementing e-learning. Moreover, different ways of increasing students’ interests and helpful references to ways of helping online students regulate their learning are also provided. The implementation of e-learning and the redesign of the materials for online courses is imperative. We expect that our experiences can provide insight for teachers in their online course design.

References


Who is responsible for E-Learning Success in Higher Education? A Stakeholders' Analysis

Nicole Wagner, Khaled Hassanein and Milena Head
DeGroote School of Business, McMaster University, Canada // wagnernm@mcmaster.ca // hassank@mcmaster.ca // headm@mcmaster.ca

ABSTRACT
Successful implementation of e-learning is dependent on the extent to which the needs and concerns of the stakeholder groups involved are addressed. This paper discusses e-learning, describes the needs and concerns of the various stakeholder groups, and derives a Stakeholders' Responsibility Matrix to summarize the responsibilities of each stakeholder group. Fulfilling the responsibilities described in the Stakeholders’ Responsibility Matrix will address the needs and concerns of each stakeholder groups, thereby encouraging the success of e-learning in higher education.

Keywords
E-learning, Higher education, Stakeholder analysis

Introduction
The environment of higher education is evolving. Rising costs, shrinking budgets, and an increasing need for distance education (New Media Consortium, 2007) are causing educational institutions to reexamine the way that education is delivered. In response to this changing environment, e-learning is being implemented more and more frequently in higher education, creating new and exciting opportunities for both educational institutions and students.

E-learning, or electronic learning, has been defined a number of different ways in the literature. In general, e-learning is the expression broadly used to describe “instructional content or learning experience delivered or enabled by electronic technologies” (Ong, Lai and Wang, 2004, page 1). Some definitions of e-learning are more restrictive than this one, for example limiting e-learning to content delivery via the Internet (Jones, 2003). The broader definition, which will be used for the purposes of this article, can include the use of the Internet, intranets/extranets, audio- and videotape, satellite broadcast, interactive TV, and CD-ROM, not only for content delivery, but also for interaction among participants (Industry Canada, 2001). More recently, this definition can be further expanded to include mobile and wireless learning applications (Kinshuk, Suonen, Sutinen, and Goh, 2003; Lehner, Nösekabel and Lehmann, 2003).

The e-learning models of higher education today find their roots in conventional distance education. Initially introduced to allow individuals in remote and rural areas to gain access to higher education, distance learning has evolved significantly over time. Technological advancement has been the major inspiration for change, beginning with the integration of radio broadcasting in the 1920’s (Huynh, Umesh and Valachich, 2003). More recently, the advent of the Internet has enabled tremendous innovation in the delivery of post secondary education (Gunasekaran, McNeil and Shaul, 2002; Teo and Gay, 2006). As time goes by, more and more people gain access to the Internet, the cost of computer ownership decreases, and overall computer literacy increases (Huynh et al., 2003). These trends provide educational institutions an ideal channel for the delivery of educational content.

Dimensions of E-Learning
The extent of e-learning technology use in course delivery varies widely. The variations in the configuration of e-learning offerings can be described through a number of attributes, as listed in Table 1 below. These attributes can be classified into the dimensions of synchronicity, location, independence, and mode. An e-learning course component can be described by indicating which one of the two attribute values from each dimension is applicable.

E-learning can be synchronous (real-time) or asynchronous (flex-time). Synchronous e-learning includes technology such as video conferencing and electronic white boards (Romiszowski, 2004), requiring students to be present at the...
time of content delivery. Asynchronous applications include programmed instruction and tutorials that allow students to work through the screens at their own pace and at their own time. Most of the courses available on the Internet are based on this asynchronous model (Greenagel, 2002). Students can be involved in e-learning from distributed locations, as in distance learning, or from the same place, such as using a group support system in a classroom to work on an assignment (Gunasekaran et al., 2002). E-learning applications also differ in the levels of collaboration that they involve. Some courses are entirely independent and individual, while others incorporate some elements of group learning such as discussion forums or chat rooms. The mode of course delivery can be entirely electronic (with or without an instructor) or take a more blended approach integrating electronic and classroom delivery to varying extents. Many current e-learning offerings follow the latter mode, taking advantage of the benefits of various types of delivery (Jack and Curt, 2001).

Table 1: The Dimensions of E-Learning

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Attribute*</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchronicity</td>
<td>Asynchronous</td>
<td>content delivery occurs at a different time than receipt by the student</td>
<td>lecture module delivered via email</td>
</tr>
<tr>
<td></td>
<td>Synchronous</td>
<td>content delivery occurs at the same time as receipt by the student</td>
<td>lecture delivery via web cast</td>
</tr>
<tr>
<td>Location</td>
<td>Same place</td>
<td>students use an application at the same physical location as other students and/or the instructor</td>
<td>using a GSS to solve a problem in a classroom</td>
</tr>
<tr>
<td></td>
<td>Distributed</td>
<td>Students use an application at various physical locations, separate from other students and the instructor</td>
<td>using a GSS to solve a problem from distributed locations</td>
</tr>
<tr>
<td>Independence</td>
<td>Individual</td>
<td>students work independently from one another to complete learning tasks</td>
<td>students complete e-learning modules autonomously</td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>students work collaboratively with one another to complete learning tasks</td>
<td>students participate in discussion forums to share ideas</td>
</tr>
<tr>
<td>Mode</td>
<td>Electronically only</td>
<td>all content is delivered via technology, there is no face-to-face component</td>
<td>an electronically enabled distance learning course</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>e-learning is used to supplement traditional classroom learning</td>
<td>in class lectures are enhanced with hands-on computer exercises</td>
</tr>
</tbody>
</table>

* The definitions of these attributes are discussed in a variety of sources including (Ong et al., 2004), (Jack and Curt, 2001), and (Greenagel, 2002)

It should be noted that a single course component will consist of a single attribute value from each dimension, yet a course may contain several components, each with different attribute values. For example, some components of a course may be delivered synchronously and others asynchronously, or a course may involve some online components and some in-class components.

E-Learning Market

Given the variety of definitions of e-learning, it is difficult to estimate the size of the market. However, e-learning is believed to be the fastest growing sub-sector of the $2.3T USD global education market, with the market for online higher education expected to grow to $69B USD by 2015 (Hezel Associates, 2005).

There are many reasons for the growth of the higher education e-learning industry, both from the institutions’ and students’ perspectives. Globally, the demand for post secondary education is increasing. For example, in the United States, college enrollment among high school graduates increased from 56% in 1980 to 67% in 2003 (Morrison, 2003). In Canada, it is expected that over 70% of new jobs created will require at least some post secondary education (Industry Canada, 2001). With the limited capacity of existing classrooms at academic institutions and the prohibitive cost of building new facilities, e-learning is an attractive alternative (Werbach, 2000).

In an effort to remain competitive and maintain their market share, many traditional higher education institutions have expanded their offerings to include e-learning courses to compete with the growing number of virtual higher education institutions (Huynh et al., 2003). In doing so, they can use their brand names to expand their target market internationally in order to capitalizing on excess demand that exists in the education systems of other countries. In
addition, increasing the use of information systems within educational institutions also provides an opportunity for the organizations to reengineer their existing time and paper-intensive processes to improve their overall efficiency (Sun Microsystems, 2003).

From a student’s perspective, the knowledge economy creates a great need for continuous training and upgrading of skills (Industry Canada, 2001). E-learning makes it possible for this lifelong learning to occur as a part of the student’s everyday life, removing the need to travel to a traditional institution or be confined to a specific class schedule.

Similar to other technology applications, the success of e-learning is dependent on the extent to which it satisfies the needs and addresses the concerns of its key stakeholders. The next section of this article contains an analysis of the key stakeholders of e-learning within the context of higher education. This analysis is then used to derive a stakeholder-to-stakeholder responsibility matrix for maximizing the chances of e-learning success within institutions of higher education, which is presented in the third section.

E-Learning Stakeholders’ Motivations and Concerns

In an organizational context, a stakeholder is a constituency of an organization (Thompson and Strickland, 2001). In the same sense, the stakeholders of e-learning are those that are affected by it. While reviewing the e-learning literature during the development of this article, a list of the main stakeholder groups in the context of higher education was compiled. Each of these stakeholder groups is described in the following sections, along with their motivations to use e-learning and their concerns about it.

Students

Students are the consumers of e-learning. In the context of higher education, they are undergraduate or graduate students enrolled at a university or college.

Motivations

Students are motivated to use e-learning to gain access to higher education. For some, it may be a component of a traditional course, while for others entire courses may be entirely online. Particularly for this second group, e-learning may create access to higher education that they would not have otherwise because of geographic or time constraints (Huynh et al., 2003; Kabassi and Virvou, 2004).

Concerns

E-learning presents an entirely new learning environment for students, thus requiring a different skill set to be successful (Romiszowski, 2004). Critical thinking, research, and evaluation skills are growing in importance as students have increasing volumes of information from a variety of sources to sort through (New Media Consortium, 2007). Also, particularly in courses that are entirely electronic, students are much more independent than in the traditional setting. This requires that they be highly motivated and committed to learning (Huynh et al., 2003), with less social interaction with peers or an instructor. Students in online courses tend to do as well as those in classrooms, but there is higher incidence of withdrawal or incomplete grades (Zhang, Zhou and Briggs, 2006).

E-learning by its very nature requires a certain level of technical sophistication. This becomes less of an issue over time as computer literacy increases. For example, in Canada, increasing proportions of young adults are going to university with enrolment of students from 18 to 24 years of age rising at a faster rate than the increase in total university enrolment (Statistics Canada, 2005). The term “digital natives” has been used to describe the under 40 generation (Prensky, 2001; Prensky, 2006). Having grown up with increased use of electronics such as television and video games, and a corresponding decrease in reading, this group learns differently than older age groups. They tend to have “a more fragmented sense of time, a reduced attention span” (Woodill, 2004, page 11) and as a result are usually disappointed and bored with mundane e-learning applications. Simulations and digital game-based learning...
may be better suited for this group. Research has linked higher levels of interactivity and learner control with increased student satisfaction in e-learning (Zhang et al., 2006).

**Instructors**

In e-learning, as in traditional classroom learning, instructors guide the educational experiences of students. Depending on the mode of e-learning delivery, instructors may or may not have face-to-face interaction with their students.

**Motivations**

Instructors may be motivated to use e-learning in their courses for a variety of reasons. For example, they may be encouraged or pressured by their institutions; they may wish to reach a broader audience of students; or they may have an interest in the benefits of technology mediated learning.

**Concerns**

E-learning technologies bring as much change to instructors as they do to students, again requiring a new set of skills for success (Jones, 2003). In the e-learning environment, instructors shift from being the primary source of students’ knowledge to being the manager of the students’ knowledge resources (Romiszowski, 2004). For example, in a traditional classroom scenario, the instructor delivers the content to the class and responds to their questions. In contrast, in a technology only asynchronous e-learning environment, the instructor is more of a coordinator of the content, which students then pursue at their own pace (Teo and Gay, 2006). Thus, the skills that are most important for an instructor to possess may depend on the e-learning attributes of their course.

E-learning requires technical sophistication from instructors as well as students (Jones, 2003). Course administration may require instructors to learn new software applications. Especially in cases where instructors are also the content creators, use of new technology may be extensive. Studies have shown that the main challenges of technical support for e-learning initiatives include lack of knowledge of how to alter instructional design to be effective for courses with technology and lack of confidence in using these applications to teach (Arabasz and Baker, 2003).

Instructors may also be concerned with the acceptance of e-learning tools among their students. Studies have found that perceived usefulness and perceived enjoyment are very important for the adoption of e-learning applications by students (for example: Mahmood, Dahlan, Ramayah, Karia and Asaari, 2005; Lee, Cheung and Chen, 2005). In order to increase perceived usefulness and enjoyment, instructors should vary the types of content, create fun, provide immediate feedback, and encourage interaction to increase acceptance (Lee et al., 2005).

The amount of time that it takes instructors to create and administer e-learning courses is another important consideration. While some promote that the delivery of e-learning courses is less labour intensive, a 2003 study found that faculty and support staff spent almost twice as many hours providing online versions of courses compared to traditional delivery (Doughty, Spector and Yonai, 2003). Unless incentives are provided to encourage instructors to use e-learning technology, resistance to additional workload is likely to occur.

**Educational Institutions**

Educational institutions, in the context of higher education, include colleges and universities. In addition to the traditional list of postsecondary institutions, the rise in popularity of e-learning has lead to the creation of new, online only educational institutions.

**Motivations**

Educational institutions integrate technology into classrooms to facilitate lecture delivery and create new technology mediated learning opportunities for students. They provide distance learning, including e-learning, to create access to
a larger pool of students. As e-learning becomes more widely accepted and more courses are offered online, geographic boundaries between institutions and students are removed (Young, 2001).

Concerns

Often, budgetary restriction is a primary issue for institutions (Huynh et al., 2003). Tight budgets make it difficult to implement broad, campus-wide e-learning solutions. There is a tendency for individual departments to implement their own solutions, which may not be consistent with the rest of the institution (Sun Microsystems, 2003). This reduces the potential for cross-departmental efficiencies, and can make the process more complicated for faculty, staff, and students, particularly if they are involved with more than one department.

Depending on the technological infrastructure in place at an institution, the implementation of e-learning courses can involve very costly technology upgrades (Weller, 2004). E-learning systems require several components including sufficient bandwidth, course management systems, technology equipped classrooms, and adequate computer facilities for student use (Arabasz and Baker, 2003). This increase in technology generally requires a corresponding increase in support staff as well (Young, 2001).

An important consideration for institutions is how the effectiveness of e-learning offerings will be assessed. Often measurement is based on return on investment (ROI) in the technology infrastructure and course content development. While this measure is certainly relevant, it is also vital to consider effectiveness in terms of learning outcomes (Romiszowski, 2004). An e-learning exercise can only be considered effective if learning took place. The tendency of organizations to focus on ROI can encourage cheaper program development, at the expense of learning effectiveness (Weller, 2004).

Resistance from faculty is another important concern for institutions. Many faculty members firmly believe that e-learning is inferior to face-to-face instruction (Huynh et al., 2003). Studies have shown, however, that there is no significant difference between the performance of students in the two methods (Huynh et al., 2003). The additional time required to administer e-learning courses, discussed above, may also contribute to resistance from faculty.

Acceptance of online education by employers is also a significant concern for institutions. If employers are less likely to hire students with online degrees, then students will be less likely to enroll in those degrees. Thus it is in the institutions best interest to encourage the acceptance of this form of education among potential employers.

Content Providers

In the higher education context, online course content may be created by instructors or acquired from external sources. The growth in e-learning has created a market for commercialized educational content creators, particularly for more introductory courses that are offered consistently at multiple institutions.

Motivations

Whether the content provider is the instructor or an external source, their motivation is to provide content modules that will result in effective learning. Commercial content providers are motivated by profit to develop content modules that are flexible enough to be readily utilized across institutions with minimal adaptation efforts.

Concerns

The main concern for content providers in e-learning tends to be intellectual capital rights (Huynh et al., 2003). Independent content providers in particular, need to ensure their retention of copy rights in order to sell their product to multiple customers.
Technology standards are another relevant concern for this stakeholder group (Teo and Gay, 2006). Content should be created in a format that will allow its utilization across various e-learning technology platforms. Failure to do so would restrict their potential target market. It is equally important to make certain that the content provided is consistent with the learning methodologies in use at various institutions and thus being more likely to result in successful learning (Greengal, 2002). Learning can be impacted by the type of content, the learning environment, and even the characteristics of each learner (Zhang et al., 2006). E-learning content providers need to take this into consideration when developing content.

**Technology Providers**

Technology providers develop the technology that enables e-learning delivery. This category consists of a broad range of services, from the facilitation of individual distance learning courses, to complete Learning Management Systems (LMS) provided by companies such as Blackboard.

**Motivations**

Similar to content providers, technology providers are motivated to provide learning environments that will result in effective learning for students.

**Concerns**

Technology standards are an important consideration for this stakeholder group as well. Since educational institutions often have different solutions implemented by various departments, adherence to common standards facilitates interoperability (Young, 2001; Friesen, 2005). Constant evolution in hardware and consumer expectations creates pressure for technology providers to rush to market with new product offerings (Huynh et al., 2003). In order for these businesses to be sustainable, the cost of pursuing this constant innovation must be controlled (Dalziel, 2003).

Many industry experts attribute the shortcomings of e-learning to technological issues (Woodill, 2004). It is argued that many products are not developed on proven educational principles and thus do not take the different ways that people learn into consideration (Woodill, 2004). Similar to content providers, technology providers should provide provisions for personalizing the learning experience based on the context of learning and the characteristics of the student.

**Accreditation Bodies**

Accreditation bodies are organizations that assess the quality of education institutions offerings. Those institutions meeting the minimum requirements will be accredited, providing them a level of credibility that non-accredited institutions will not possess.

**Motivations**

As the proportion of education delivered by electronic means grows, it is increasingly important for accreditation bodies to encompass e-learning in their standards. Neglecting to do so will limit the relevance of their accreditation since it will only be relevant to the traditional education component of educational institutions’ offerings.

**Concerns**

The growth of e-learning presents new challenges for accreditation bodies. As the number of learning institution grows in an attempt to capitalize on the excess demand for higher education, accreditation bodies have an increasing number of institutions seeking their approval. This increase in volume of work is combined with a change in the nature of the work that these bodies do. The Council for Higher Education Accreditation (CHEA) in the United


States defines distance learning as educational or instructional activity that is delivered electronically to students at a distance (CHEA, 2002b). By this definition, all distance learning (including e-learning) is subject to the same accreditation and securitization.

In order for e-learning courses and degrees to be taken seriously within higher education and by employers, it is very important that proper accreditation processes exist for their evaluation. Accreditation bodies recognize that there are unique considerations for evaluating e-learning. The three major concerns are: alternative design of instruction, alternative providers of higher education, and expanded focus on training (CHEA, 2002a). These areas involve issues such as the appropriateness of subject matter, the technology platform and teaching methodology chosen, technical support for faculty and students, and many more.

Employers

Employers, in this context, are those organizations that will potentially hire graduates of higher education institutions. Often, there is a tendency for employers to view online education from reputable traditional institutions in a more positive light; however the acceptance of online degrees in general is increasing (Chaney, 2002). This is a positive trend for e-learning in general and for completely online educational institutions in particular.

Motivations

Employers are increasingly motivated to consider e-learning as a higher education alternative. Denying the value of e-learning will restrict their pool of potential hires. It will also limit the availability of courses and professional development activities that their employees may participate in.

Since many students pursue higher education for the purpose of beginning or advancing their careers, a lack of support for e-learning by employers could deter employees from pursuing their coursework through electronic means, thereby restricting their opportunities.

Concerns

One issue that employers have with e-learning is the decreased interpersonal interaction inherent in many of these courses. Employers typically rank technical skills and expertise from 6 to 8 on a scale of 10, and rank interpersonal skills to be of higher importance (Gunasekaran et al., 2002). Some feel that while e-learning may be suitable for delivering content, it may not be capable of developing these interpersonal skills that employers value so highly.

E-Learning Stakeholders’ Responsibility Matrix

The various stakeholders in higher education e-learning interact with one another in a variety of ways. The success of e-learning is thus dependant on the cooperation of all of those stakeholder groups. Consequently, each stakeholder group has responsibilities towards the other stakeholders to help fulfill their motivations and address their concerns. Taking into consideration the needs and concerns of each stakeholder group as discussed above, we developed a Stakeholder Responsibility Matrix, as shown in Table 2 (below). This matrix outlines the actions that each stakeholder group should take in order to address the motivations/needs and concerns of the other groups. In populating this matrix, we relied on our knowledge of the e-learning literature in addition to the experience of the first author as a former MBA student and current Ph.D. student in an environment where e-learning was available and utilized. We also relied on the experience of the second and third authors as experienced professors who have employed various forms of e-learning within their undergraduate and graduate courses over the past 9 years. Additionally, the third author has direct instructional design experience working with external content provider consultants and various accreditation bodies. The following process was followed: (i) the first author proposed an initial matrix based on extensive literature review and personal student experience; (ii) the second and third authors examined the proposed matrix independently and proposed additions and modifications; and (iii) the revised matrices (from the second and third authors) were compared for similarity and any discrepancies were discussed as a group to derive the finalized matrix. This feedback methodology using multiple expert judges is advocated by Straub (1989).
The matrix reads from left to right, indicating the responsibilities of the group down the first column to each group listed across the first row. For example, consider the obligations between Institutions and Instructors. Institutions responsibilities to Instructors include providing training in both instructional design and technology use, providing technical support, providing incentives to incorporate e-learning, and enforcing standardization to promote better technical performance. Conversely, Instructors responsibilities to Institutions centre on using the e-learning technologies available to them according to the institutions policies and standards. Fulfilling the responsibilities described in the Stakeholders’ Responsibility Matrix will address the needs and concerns of each stakeholder group, thereby encouraging the success of e-learning in higher education.

Conclusion

E-learning is a large and growing market with great potential in higher education. In order to maximize this potential, e-learning implementations should endeavor to satisfy the needs and concerns of all stakeholder groups as much as possible. The Stakeholders’ analysis undertaken in this paper and culminating in the Stakeholders’ Responsibility Matrix is a step in that direction.

As shown in our e-learning Stakeholders’ Responsibility Matrix, each stakeholder group has an important role to play while working together towards the common goal of enhancing the overall learning experience. Students and Instructors should participate as proactively as possible, provide feedback to improve future experiences, and communicate the learning possibilities that e-learning creates. Institutions should provide the technical infrastructure and support needed to enable comprehensive solutions. Content and Technology Providers should provide high quality, interoperable solutions that consider learning principles. Accreditation Bodies should provide and enforce clear guidelines for this new form of learning delivery. Employers need to recognize the validity of this form of education and work with other stakeholders to ensure that graduates meet the needs of the job market.

Institutions of higher education could utilize the stakeholders’ responsibility matrix presented in this paper as a starting point when undertaking a new e-learning initiative. The stakeholders involved and their associated responsibilities could then be adapted to the nature of the particular initiative at hand. As such, the matrix will help institutions to identify the appropriate stakeholders’ and develop a set of expectations for each.

Since the presented framework involves the cooperation of each stakeholder group, its implementation would entail communication between groups to ensure that the responsibilities of each group are clear. To this effect, when institutions undertake a substantial e-learning initiative they should strive to involve a cross-functional team with representation from each relevant stakeholder group. This will ensure their specific needs are addressed during development and will help to facilitate buy-in during implementation. Successful implementation also requires a project champion, who will communicate the responsibilities and the importance of cooperation to each group. It is suggested that leadership from the highest level of the institution is needed to see the opportunities available and bring them to reality (New Media Consortium, 2007). Through the effective dissemination of information, those involved in e-learning can be made aware of how they fit into the complete picture, and the importance of their specific roles in e-learning implementation success.

The framework presented in this paper is derived from the motivations/needs and concerns of stakeholder groups as noted in the literature. Future research should be conducted to validate this framework across various institution types, educational programs, and cultural settings. Since the application of the framework at an institution requires the coordination of many stakeholder groups, a case study methodology may be the most appropriate to study the application of the framework. This methodology would allow for in depth study of the success of a particular application in light of the levels of cooperation achieved according to the Stakeholders’ Responsibility Matrix.

Returning to the question posed in the title of this paper: “who is responsible for e-learning success in higher education?”, we have shown the answer to be a shared responsibility between the various e-learning stakeholders. When all stakeholders fulfill their responsibilities to create effective and meaningful e-learning experiences, positive outcomes extend beyond success in specific courses and programs to facilitate lifelong learning and discovery.
<table>
<thead>
<tr>
<th>E-Learning Stakeholders</th>
<th>Responsibility Matrix*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
<td>• Participate in collaborative exercises to enhance learning</td>
</tr>
<tr>
<td></td>
<td>• Encourage and share experiences among peers</td>
</tr>
<tr>
<td><strong>Instructor</strong></td>
<td>• Engage proactively in exercises</td>
</tr>
<tr>
<td></td>
<td>• Provide feedback regarding overall effectiveness</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td>• Use e-learning technologies according to institutional policies</td>
</tr>
<tr>
<td></td>
<td>• Encourage the use of e-learning services and standards</td>
</tr>
<tr>
<td><strong>Content Provider</strong></td>
<td>• Provide content that meets institutional needs</td>
</tr>
<tr>
<td></td>
<td>• Comply with learning &amp; usability standards</td>
</tr>
<tr>
<td><strong>Technology Provider</strong></td>
<td>• Comply with standards for interoperability</td>
</tr>
<tr>
<td><strong>Accreditation Body</strong></td>
<td>• Promote the validity of e-learning</td>
</tr>
<tr>
<td><strong>Employer</strong></td>
<td>• Ensure that standards provide appropriate measures</td>
</tr>
</tbody>
</table>

* Matrix generated by the authors according to the procedure outlined in Section 3.
References


Chaney, E.G. (2002). Pharmaceutical Employers' Perceptions of Employees or Applicants with E-Degrees or Online Coursework, Dissertation, Indiana State University, USA.


A Survey on ICT Usage and the Perceptions of Social Studies Teachers in Turkey

Yasemin Gulbahar
Faculty of Education, Baskent University, Turkey // Tel: +90 312 2341010/1076 // Fax: +90 312 2341174 // gulbahar@baskent.edu.tr

Ismail Guven
Faculty of Educational Sciences, Ankara University, Turkey // Tel: +90 312 3633350/5203 // Fax: +90 312 3636145 // iguven@mailcity.com

ABSTRACT

Turkey has been undertaking many projects to integrate Information and Communication Technology (ICT) sources into practice in the teaching-learning process in educational institutions. This research study sheds light on the use of ICT tools in primary schools in the social studies subject area, by considering various variables which affect the success of the implementation of the use of these tools. A survey was completed by 326 teachers who teach fourth and fifth grade at primary level. The results showed that although teachers are willing to use ICT resources and are aware of the existing potential, they are facing problems in relation to accessibility to ICT resources and lack of in-service training opportunities.

Keywords

Information and communication technology, Social studies teachers

Introduction

Throughout the world, many countries have introduced Information and Communication Technologies (ICT) into schools via different courses of action. Their use is also underlined by OECD (2001) as a necessity for improving quality in teaching and learning. The Ministry of National Education (MNE) in Turkey has also made huge investments in the hope of attaining the goal of improving the quality of education through enriching the learning environment with the help of educational software and technologies. Integrating ICT training into all levels of primary education and providing each student with access to ICT equipment and information sources were also among the objectives of the MNE. In 2001, 2837 ICT classrooms were established. The distribution of educational software purchased for these schools was also completed in the same year. ICT classrooms are equipped with computers, printers, instructional software, electronic references, video players, overhead projectors and TV. The policy makers in Turkey expected that the introduction of ICT into formal education settings would improve the academic performance of teachers by encouraging them to improve their ability to use and apply technology and software in their jobs. Programs have been organized for teachers to access to ICT in every circumstance (MNE, 2003). Furthermore, in-service training opportunities for many teachers in different subject areas have been provided. It was hoped that teachers’ use of technology in education would improve educational outcomes, increase technological skills and reduce anxiety when preparing lessons. Technology usage is an important indicator of their preparedness to carry out the obligations of daily lessons. In fact, Woodrow (1992) asserts that any successful transformation in educational practice requires the development of positive user attitude toward new technology. The development of teachers’ positive attitudes toward ICT is very significant factor not only for increasing computer integration but also for avoiding teachers’ resistance to ICT use (Watson, 1998).

The Need for ICT Integration in Schools

ICT integration in schools is needed in order to accomplish many objectives and improve the quality of lessons in all subject areas as well as social studies. ICT increasingly pervades various aspects of our daily lives like work, business, teaching, learning, leisure and health. Since ICT leads all processes based on information, every individual in a society should become technology competent. Thus, all schools have to be equipped with the necessary ICT in order to provide the next generations with the needed tools and resources for access and use and to attain the expected skills. Norris, Sullivan and Poirot (2003) point out the importance of accessibility as: “…teachers’ use of technology for curricular purposes is almost exclusively a function of their access to that technology” (p. 25). Merely
providing schools with hardware, software and in-service training is not enough. Any in-service training needs follow-up support, peer coaching and peer dialogue to ensure successful utilization of new technologies. There must be active involvement of the teachers concerned in the whole change process so that there is the element of “ownership” of the innovation.

Just filling schools with the necessary ICT neither improves the quality of instruction nor creates more effective learning environments. However, embracing a broader vision and philosophy, schools should revise present teaching programs, practices and resources, and ICT should be integrated into all levels of an educational system from classrooms to ministries for use in management, teaching and learning activities. Thus, “Teachers must receive adequate ongoing training, technology use must be matched to curriculum’s philosophy and theory of learning, and adequate numbers of computers must be conveniently located within the classroom” (Al-Bataineh & Brooks (2003), p. 479). As also concluded by Kington, Harris and Leask (2002) “…it is not necessarily the technology that has to be innovative, but the approach to teaching and learning must be” (p. 35).

**Teachers’ ICT Usage**

The integration of information and communication technologies can help revitalize teachers and students. This can help to improve and develop the quality of education by providing curricular support in difficult subject areas. To achieve these objectives, teachers need to be involved in collaborative projects and development of intervention change strategies, which would include teaching partnerships with ICT as a tool. Teachers’ attitudes are major predictors of the use of new technologies in instructional settings. Teachers’ attitudes toward ICT shape not only their own ICT experiences, but also the experiences of the students they teach. According to Zhao and Cziko (2001) three conditions are necessary for teachers to introduce ICT into their classrooms: teachers should believe in the effectiveness of technology, teachers should believe that the use of technology will not cause any disturbances, and finally teachers should believe that they have control over technology (p. 27). Demetriadis et al. (2003) reached similar conclusions in their research study: “Training efforts are generally welcomed by teachers but consistent support and extensive training is necessary in order for them to consider themselves able to integrate ICT in their teaching methodologies” (p. 35). According to Rogers (1995) one of the major factors affecting people’s attitudes toward a new technology is related to the features of the technology itself. Rogers points out five basic features of technology that affect its acceptance and subsequent adoption: relative advantage, compatibility, complexity, observability, and trialability. Thus, a new technology will be increasingly diffused if potential adopters perceive that the innovation: (1) has an advantage over previous innovations; (2) is compatible with existing practices, (3) is not complex to understand and use, (4) shows observable results, and (5) can be experimented with on a limited basis before adoption.

Preparing students for real life in our technological and diverse world requires that teachers embed ICT in significant learning experiences (Braun & Kraft, 1995). However, research studies show that most teachers do not make use of the potential of ICT to contribute to the quality of learning environments, although they value this potential quite significantly (Smeets, 2005). Harris (2002) conducted case studies in three primary and three secondary schools, which focused on innovative pedagogical practices involving ICT. Harris (2002) concludes that the benefits of ICT will be gained “…when confident teachers are willing to explore new opportunities for changing their classroom practices by using ICT” (p. 458). As a consequence, the use of ICT will not only enhance learning environments but also prepare next generation for future lives and careers (Wheeler, 2001).

**Research Questions**

The purpose of this study was to explore ICT usage, factors that support the use, barriers that hinder the use, and self-perceptions of efficacy and level of expertise, as well as the relationship of variables by looking at the social studies teachers in selected primary schools in Turkey. In order to shed light on these topics, this research study mainly focused on the following nine questions.

1. Which ICT resources (software, instructional tools and materials) do social studies teacher’s use?
2. What are teachers preferred methods for professional development?
3. What are the incentives that encourage social studies teachers’ technology usage?
4. What are teachers’ perceptions of self-efficacy in relation to ICT usage?
5. What are the barriers social studies teachers face during technology usage in the teaching-learning process?
6. Is there any relationship between awareness and self-rated expertise level of teachers and, between self-perception of efficacy and self-rated expertise level of teachers?
7. Is there any relationship between teachers’ computer related tools usage in the classroom and self-perceptions of efficacy?
8. Is there any relationship between teachers’ computer related tools usage in the classroom and level of expertise?
9. Is there any relationship between having a computer at home and the expertise level of teachers?

**Method**

Convenience sampling was used to reach the participants in this study. The participants for this study were 326 social studies teachers from fourth and fifth grade of various primary schools, who voluntarily participated in the study.

The “Information and Communication Technology Usage Survey” (α = 0.84) developed by the researchers, mainly based on discussions in the related literature (Iding, Crosby & Speitel (2002); Bielefeldt (2001); Haydn, Arthur & Hunt (2001); McCormick & Scrimshaw, 2001) was used to collect data for this research study. The survey was composed of five parts. The first part of the survey consisted of 24 items regarding teachers’ software use, as well as other instructional tools and materials. The purpose of this part was to find out the self-expertise level of the social studies teachers. The second part consisted of 9 items about preferences for professional development on information gathering and support. The subsequent part consisted of 8 items about factors that encourage teachers’ usage of technology. In the fourth part of the survey there were 18 items related to teachers’ perceptions of self-efficacy. Finally, the last part was composed of 19 items regarding the barriers that teachers faced during technology utilization in the teaching-learning process.

**Results**

In this part of the study, the results of the previously listed research questions are reported.

**ICT resources used by social studies teachers**

The majority of social studies teacher in this study, 98.2% have access to a computer at work and among them 88.7% have access to the Internet. Daily computer usage of social studies teachers was found to be as follows: 53.1% uses a computer for less than one hour, 30.7% uses a computer for between 1 and 3 hours, 2.8% uses a computer for between 3-5 hours and 1.5% uses a computer for more than five hours a day.

The social studies teachers specified their level of expertise on thirteen types of computer software by using a three-point likert-type scale (that is, 2=Good, 1=Average and 0=None) (α=0.93). Over fifty-four percent of the participants rated their skills as average or high at word processing, spreadsheets, presentation software, computer aided instructional software, web browsers, search engines, electronic mail, chat/forum, electronic encyclopedias and instructional films.

The social studies teachers indicated their usage of eleven types of instructional tools and materials by using a three-point likert-type scale (2=Often, 1=Sometimes and 0=None) (α=0.81). The preferred instructional tools according to usage rate are as follows: board, printed materials, overhead projector, television/video, radio cassette recorder, multimedia computer and slide projector.

**Teachers preferred methods for professional development**

The social studies teacher’s preferences for professional development, namely accessing knowledge (5 questions) and support services (4 questions), were taken through a three-point likert type scale (2=I prefer, 1=Neutral and 0=I
don’t prefer) (α=0.66). Printed materials (99.4%), Internet resources (83.4%) and self-study (80.7%) and participation in seminars and workshops (79.1%) were the most favored knowledge resources for professional development. On the other hand, the majority of teachers favored every kind of support service: experienced teachers (96.9%), colleagues in the same field (87.7%) and technical support group within the school (81.0%).

The incentives that encourage social studies teachers’ technology usage

The participants used a three-point likert-type scale (i.e. 2=Important, 1=Neutral and 0=Not Important) to rate their level of importance on 8 statements about incentives for adoption (Table-1). All the statements were rated as important incentives by over 80% of the teachers. (α=0.73).

<table>
<thead>
<tr>
<th>Factors Encourage Technology Usage</th>
<th>Important (%)</th>
<th>Neutral (%)</th>
<th>Not Important (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rewarding the technology usage efforts of teachers in instructional activities</td>
<td>80,4</td>
<td>8,3</td>
<td>10,4</td>
</tr>
<tr>
<td>Investments of the institution on infrastructure of instructional technologies</td>
<td>96,6</td>
<td>3,1</td>
<td>0</td>
</tr>
<tr>
<td>Investments of the institution on in-service education programs for instructional technologies</td>
<td>90,8</td>
<td>8,9</td>
<td>0</td>
</tr>
<tr>
<td>Investments of the institution on the support services of instructional technologies</td>
<td>84,7</td>
<td>12,9</td>
<td>0</td>
</tr>
<tr>
<td>Developing the policies and plans for diffusion of the instructional technologies</td>
<td>89,0</td>
<td>9,5</td>
<td>1,2</td>
</tr>
<tr>
<td>Providing support for the projects towards the expansion of instructional materials</td>
<td>92,0</td>
<td>5,8</td>
<td>1,8</td>
</tr>
<tr>
<td>Carrying out the studies for integration of technology into curriculum</td>
<td>88,7</td>
<td>9,2</td>
<td>1,8</td>
</tr>
<tr>
<td>Reducing work load to provide opportunities to teachers for developing instructional materials</td>
<td>91,7</td>
<td>4,9</td>
<td>3,4</td>
</tr>
</tbody>
</table>

Teachers’ perceptions of self-efficacy in relation to ICT usage

The participants used a three-point likert-type scale (i.e. 2=Agree, 1=Neutral and 0=Disagree) to specify their perceptions on 18 statements about using computers and instructional technologies (α=0.62) (Table-2). The results showed that teachers believe that technology will bring to them advantages, but they lack the basic skills of computer usage, and they also feel that their skills are lacking for other technologies which could also be used as an aid in the classroom.

<table>
<thead>
<tr>
<th>Perceptions</th>
<th>Agree (%)</th>
<th>Neutral (%)</th>
<th>Disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t use computers as much as other resources (books, overhead projectors etc.) for instructional purposes.</td>
<td>42,0</td>
<td>5,5</td>
<td>52,1</td>
</tr>
<tr>
<td>I know what to do for using computers in instructional environments.</td>
<td>48,5</td>
<td>29,8</td>
<td>20,6</td>
</tr>
<tr>
<td>I am aware of the opportunities that computers offer.</td>
<td>28,5</td>
<td>34,0</td>
<td>33,1</td>
</tr>
<tr>
<td>I can answer any question my students ask about computers.</td>
<td>19,3</td>
<td>16,3</td>
<td>64,1</td>
</tr>
<tr>
<td>I am not sure that I am computer-literate for use computers in my classes.</td>
<td>23,3</td>
<td>27,0</td>
<td>48,8</td>
</tr>
<tr>
<td>I don’t want to use computers.</td>
<td>55,5</td>
<td>13,5</td>
<td>27,3</td>
</tr>
<tr>
<td>I think that I can use instructional technologies in class activities more effectively day by day.</td>
<td>70,9</td>
<td>18,7</td>
<td>10,1</td>
</tr>
<tr>
<td>I believe that tools like e-mail, forum and chat will make communication with my colleagues and students easier.</td>
<td>90,5</td>
<td>6,4</td>
<td>3,1</td>
</tr>
<tr>
<td>I think that technology supported teaching makes learning more effective.</td>
<td>95,7</td>
<td>2,1</td>
<td>1,8</td>
</tr>
</tbody>
</table>
I think the use of instructional technologies increases the interest of students toward courses. 97,5 2,1 0
I think the use of instructional technologies increases the quality of courses. 94,8 4,6 0,3
I think that usage of instructional technologies makes it easier to prepare course materials (assignments, handouts etc.). 90,8 7,1 1,8
It is hard for me to explain the use of computer applications to my students. 37,4 21,8 38,3
I can handle different learning preferences of my students having different learning styles by using instructional technologies. 25,8 44,8 29,1
I think technology makes effective use of class time. 77,9 17,8 4,0
I think using instructional technologies makes me more productive as a teacher. 94,2 4,9 0
I think that using technology makes it easier to reach instructional resources. 95,4 0,6 3,7
I don’t prefer to be assessed about my instructional technology based applications by any other professionals. 54,2 19,3 24,2

**Barriers social studies teachers face during technology usage**

The participants used a three-point likert-type scale (i.e. 2=Agree, 1=Neutral and 0=Disagree) to rate their level of agreement on 19 statements about barriers to adoption ($\alpha=0.87$) (Table-3). Of the 19 statements, 17 have been rated by more than 50% of the teachers as major barriers to adoption of technology into the teaching-learning process. Of these 17 statements the top three are; (1) inefficiency of teachers’ technical knowledge to prepare materials based on technology, (2) inadequacy of the technology courses offered to students and (3) lack of incentives for encouraging technology usage.

<table>
<thead>
<tr>
<th>Barriers to Technology Usage</th>
<th>Agree (%)</th>
<th>Neutral (%)</th>
<th>Disagree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inefficient time to prepare materials based on technology</td>
<td>67,2</td>
<td>9,2</td>
<td>23,9</td>
</tr>
<tr>
<td>Inefficiency of teachers’ technical knowledge to prepare materials based on technology</td>
<td>94,2</td>
<td>4,3</td>
<td>1,2</td>
</tr>
<tr>
<td>Problems about accessibility to existing hardware (computer, overhead projector etc.)</td>
<td>70,6</td>
<td>11,0</td>
<td>18,1</td>
</tr>
<tr>
<td>Inefficiency of institutions computer laboratory</td>
<td>69,0</td>
<td>4,0</td>
<td>26,7</td>
</tr>
<tr>
<td>Inefficiency of institutions technical infrastructure about instructional technology</td>
<td>55,2</td>
<td>6,7</td>
<td>37,7</td>
</tr>
<tr>
<td>Inefficient number of media (printer, scanner etc.) for effective use of computers</td>
<td>62,3</td>
<td>11,0</td>
<td>26,4</td>
</tr>
<tr>
<td>Shortage of computers used by teachers</td>
<td>65,6</td>
<td>6,0</td>
<td>33,4</td>
</tr>
<tr>
<td>Absence of reward systems for encouraging technology usage</td>
<td>73,9</td>
<td>15,0</td>
<td>10,7</td>
</tr>
<tr>
<td>Poor technical and physical infrastructure of learning environments.</td>
<td>69,9</td>
<td>8,6</td>
<td>21,2</td>
</tr>
<tr>
<td>Inadequacy of computers used by learners</td>
<td>69,3</td>
<td>7,1</td>
<td>23,3</td>
</tr>
<tr>
<td>Inefficiency of guidance and support by administration</td>
<td>68,7</td>
<td>8,9</td>
<td>22,1</td>
</tr>
<tr>
<td>Insufficiency of financial resources for technology integration</td>
<td>69,6</td>
<td>16,9</td>
<td>13,2</td>
</tr>
<tr>
<td>Inefficiency of instructional software/electronic resources</td>
<td>62,6</td>
<td>13,8</td>
<td>22,7</td>
</tr>
<tr>
<td>Scarcity in resources on technology for attaining information</td>
<td>50,3</td>
<td>18,7</td>
<td>30,7</td>
</tr>
<tr>
<td>Deficiency in professional development opportunities for gaining knowledge and skill</td>
<td>65,3</td>
<td>16,3</td>
<td>18,1</td>
</tr>
<tr>
<td>Deficiency in support services in material development/technology usage</td>
<td>47,5</td>
<td>21,8</td>
<td>29,1</td>
</tr>
<tr>
<td>Lack of interest of teachers in technology usage</td>
<td>66,6</td>
<td>12,9</td>
<td>20,2</td>
</tr>
<tr>
<td>Difficulties of improper teaching methods for technology usage</td>
<td>54,6</td>
<td>34,4</td>
<td>6,1</td>
</tr>
<tr>
<td>Inadequacy of the courses of technology offered to teachers</td>
<td>80,1</td>
<td>9,5</td>
<td>10,1</td>
</tr>
</tbody>
</table>

**The relationship between awareness and self-rated expertise level of teachers and between self-perception of efficacy and self-rated expertise level of teachers**

Correlation analysis was conducted to determine if there are any relationships between awareness and self-rated expertise level of teachers and, between self-perception of efficacy and self-rated expertise level of teachers. The
results indicated that there is no significant relationship between teachers’ awareness and their self-rated expertise level. However, a significant correlation between teachers’ self-perception of efficacy and teachers’ self-rated expertise has been identified.

Table 4. Correlational Analysis between self-perception of efficacy and self-rated expertise level of teachers

<table>
<thead>
<tr>
<th>Computer Software Usage</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
<th>Self-Perceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Software Usage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>,552**</td>
<td>316</td>
<td>,000</td>
</tr>
<tr>
<td></td>
<td>,000</td>
<td>,552**</td>
<td>288</td>
<td>,000</td>
</tr>
<tr>
<td>Self-Perceptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Perceptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>,552**</td>
<td>,000</td>
<td>288</td>
<td>,000</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

The relationship between teachers’ use of computer related tools in the classroom and self-perceptions of efficacy

To determine the proportion of variance in the attitudes of teachers toward ICT in education that could be explained by the selected independent variables, simple correlations were performed. Simple correlations (using Pearson and Spearman analyses) were first performed to identify independent variables that individually correlate with self-perception of efficacy and related tools usage in the classroom for each of the four tools: multimedia computer, computer-aided educational software, computer-projector system and the Internet/Web Environment.

Table 5. Correlational Analysis between Relationships of teachers’ computer related tools usage in classroom and self-perceptions of efficacy

<table>
<thead>
<tr>
<th>Computer Software Usage</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
<th>Self-Perception of Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Software Usage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>,553**</td>
<td>326</td>
<td>,000</td>
</tr>
<tr>
<td></td>
<td>,000</td>
<td>,553**</td>
<td>288</td>
<td>,000</td>
</tr>
<tr>
<td>Self-Perception of Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Perception of Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>,553**</td>
<td>,000</td>
<td>288</td>
<td>,000</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Follow-up tests were conducted to evaluate pairwise differences among the means. The results of these tests, as well as means and standard deviations for multimedia computer, computer-aided educational software, computer-projector system and the Internet/Web environment are reported in Table-5. The results indicated that the groups who sometimes and frequently use multimedia computer and computer-projector system in the classroom have a higher self-perception of efficacy than the group that never uses them. Moreover, the groups that frequently use educational software and the Internet/Web environment have a higher perception of efficacy than the one who never used them. In other words, teachers who have high perception of efficacy tend to use computer related tools in the classroom more frequently than the others.

The relationship between teachers’ use of computer related tools in the classroom and level of expertise

The expertise level of teachers in a classroom motivates teachers’ use of ICT more effectively. The expertise level of teachers has been analyzed here. Correlation analysis was conducted to determine if there is any relationship between teachers’ use of computer related tools in the classroom and the expertise level of teachers. A one-way analysis of variance was conducted to evaluate the relationship between the level of expertise and computer related tools usage of social studies teachers in the classroom for each of the four tools: Multimedia Computer, Computer-Aided Educational Software, Computer-Projector System and the Internet/Web Environment.
Follow-up tests were conducted to evaluate pairwise differences among the means. The results of these tests, as well as means and standard deviations for multimedia computer, educational software, computer-projector system and the Internet/Web environment are reported in Table-6. The results indicated that the groups that sometimes and frequently use computer related tools in the classroom have a higher level of expertise than the groups that never use them. In other words, teachers who have a high level of expertise tend to use computer related tools in the classroom more frequently than the others.

Table 6. Correlational Analysis between Relationships of teachers’ computer related tools usage in classroom and level of expertise

<table>
<thead>
<tr>
<th>Computer Software Usage</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.</td>
<td>323</td>
</tr>
<tr>
<td>Level of expertise</td>
<td>.552**</td>
<td>.000</td>
<td>288</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

The relationship between having a computer at home and the expertise level of teachers

An independent-samples t-test was conducted to evaluate if there is a relationship between having computer at home and the expertise level of teachers. The results showed that teachers who have computer at home have higher level of expertise than the others (t (312)=-3.891, p<0.01, Means 10.3849 vs 4.8261).

Discussion

Educators wishing to support the integration of ICT into subject teaching need to overcome the organizational and political obstacles that occur as well as certain limiting perspectives, both personal and professional, that some teachers’ may hold. There are few studies that have analyzed how subject cultures differentially affect teachers’ use of ICT. Studying teacher perspectives on ICT allows us to suggest further methods for successful integration into the core subjects. This is the first attempt to make explicit how teachers go about integrating ICT into core subject social studies in Turkey.

Rapid growth and improvement in ICT have led to the diffusion of technology in education. Studies in controlled environments suggest that the use of technology under the right circumstances improves educational outcomes, and many educators believe that a new pedagogy that incorporates technology is necessary to prepare students for work in the information age. The study investigated the perceptions and ICT usage of social studies teachers. Perceptions and skills in relation to ICT have been universally recognized as an important factor in the success of technology integration in education. Findings from this study suggest that social studies teachers understand the benefits of ICT usage in education. Social studies teachers considered computers as a viable educational tool that has the potential to bring about different improvements to their schools and classrooms. The findings of the study indicated a very strong positive correlation between teachers’ attitudes toward ICT in education and their perceptions of the advantages of the use of computers. However, teachers’ perceptions of the compatibility of ICT with their current teaching practices were not as positive. Teachers pointed out that the class time is too limited for ICT usage. Hence, the introduction of ICT innovations into education requires promoting structural, pedagogical and curricular approaches. Cultural perceptions should be taken into consideration. This conclusion points to the need for considering cultural factors in studies conducted in developing countries (Albrini, 2006).

This study examined the extent to which teachers have access to ICT in assessing the frequency with which teachers used computers for various activities. Social studies teachers mostly preferred board, printed materials, overhead projectors, television/video, radio cassette recorder, multimedia computer and slide projector for instructional aims. Teachers most frequently used computers for accessing information on the Internet, communicating electronically, doing word processing and making slide presentations. Only a few teachers reported using ICT to help them learn
school material, and less than one-fifth regularly used educational software. Some teachers reported using ICT for programming, drawing, graphics or analyzing data with spreadsheets, but this is very rare.

Printed materials (99.4%), Internet resources (83.4%), self-study (80.7%) and participating in seminars and workshops (79.1%) were the most favored knowledge resources for professional development. On the other hand, the majority of teachers favored every kind of support service: experienced teachers (96.9%), colleagues in the same field (87.7%) and technical support group within the school (81.0%). Social studies teachers focused on elective courses and other short, in-service professional development courses and workshops for professional development. Teachers pointed out the need for some sharing of experiences and discussion of new technologies and contemporary issues, so that teachers receive support in trying to keep up with new developments in ICT. In addition to longer practical work, teachers needed more resource materials such as supplementary workbooks and a resource center where they could find teaching materials and ideas.

The majority of teachers acknowledged the importance of using ICT in their own teaching. The majority of teachers also reported a lack of confidence in applying ICT in their teaching. All teachers maintained an increased enthusiasm to apply ICT in their teaching in every circumstance. Based on these results, the training course succeeded in giving the teachers enhanced skills in pedagogical and technical use of the ICT-based learning, program components and an increased motivation for using ICT.

The self-expressed feeling of social studies teachers, that they lacked the “technical knowledge to prepare materials based on technology”, showed the importance of in-service training and paralleled the result on instructional tools and materials usage. Moreover, social studies teachers appear to be unaware of possible technologies that could be helpful in the teaching processes and the majority does not use ICT. On the other hand, all kinds of professional development preferences and support service opportunities were highly rated, showing the willingness of teachers to learn and highlighting the lack of in-service training opportunities.

What are teachers’ perceptions of self-efficacy in relation to ICT usage? ICT were introduced into schools not as a means, but as an end. There were no supplementary measures to enable educators to develop positive attitudes toward the new tools and to use them. This has often resulted in ad hoc approaches to implementation. It is necessary for teachers to have the appropriate skills, knowledge and attitudes to integrate ICT into the curriculum. That is, teachers should become effective agents to be able to make use of technology in the classroom. Ultimately, teachers are the most important agents of change within the classroom arena. This result is similar to that stated by Zhao, Pugh, Sheldon and Byers (2002): “…teachers need to know the affordances and constraints of various technologies and how specific technologies might support their own teaching practices and curricular goals. They also need to know how to use technologies” (p. 511).

The results indicated that there is no significant relationship between teachers’ awareness and their self-rated expertise level. However, a strong relationship between teachers’ self-perception of efficacy and teachers’ self-rated expertise has been identified (Pearson r= 0.552, p<0.01). The lack of confidence in using ICT in teaching, as observed in this study, could be due to the fact that the in-service teacher training course was not tailored to the participants’ needs. These conditions need to have been met in order for the course to act as a positive enabling factor. The design of the in-service training should have followed guidelines in use, which states that a stepwise activity should be provided along with the combination of input, practice, reflection on practice and new input. In order to solve these problem in-service courses should be organized on the basis of introductory problem based–learning workshops, where teachers could learn and practice the method to be used during the implementation of the learning program. The results showed that although teachers are willing to use ICT resources and are aware of the existing potential, they are facing problems with accessibility to ICT resources and lack of in-service training opportunities. These two main points are also underlined by Galanouli and McNair (2001) when they state: “…schools must be supported and resourced properly, and teachers must have effective ICT training, before improvements in school-based ICT development for student-teachers can be achieved” (p. 396). It has been found that social studies teachers use computer technology, especially application software or tool applications. This shows a trend towards the use of application software in classroom teaching and learning. Thus, personal, reinforcing and enabling factors must be taken into consideration for the planning of ICT-based in-service teacher training. It is especially important to foster a feeling of ownership among the teachers towards the learning program in addition to support from colleagues and the school leaders.
As the favored incentive, the social studies teachers selected the item “Investments of the institution on infrastructure of instructional technologies”. This finding highlights the fact that access to ICT resources should be one of the primary goals to be met. Related to this topic, Zhao, Pugh, Sheldon and Byers (2002) state that; “Although in recent years there is a great progress in bringing computers and networks to schools, we found that in many schools teachers did not have easy access to either of the two infrastructures” (p. 512), which is also similar to the case in Turkey. The national programs have been of limited success not only because they were formulated in non-educational realms, but also because they were not based on research.

Social studies teachers pointed out that one of the main barriers to technology implementation is insufficiency of teachers’ technical knowledge to prepare materials based on technology. This shows us that equipping schools with ICT is not enough for attaining educational change. The introduction of ICT into education requires an equal level of innovation in other aspects of education. The inadequacy of the technology courses offered to teachers and the lack of incentives for encouraging technology are further barriers to ICT usage. Teachers’ attitudes toward computer technologies are also related to teachers’ computer competence. Teachers’ computer competence is a significant predictor of their attitudes toward computers. Teachers who have difficulty using ICT maintained that the main barriers were lack of knowledge and skills with computers that would enable them to make “informed decisions”. The results related to self-efficacy validated the findings of both inefficient use of technology due to lack of knowledge and strong belief in the potential of using technology in and out of class activities. Besides these, results once more indicated that most of the social studies teachers are computer-illiterate and they need in-service training. This finding was also paralleled by the finding that the major barrier stated by almost all of teachers was “inefficiency of teachers’ technical knowledge to prepare materials based on technology”. The findings show that social studies teachers have high awareness (Mean=15.24, SD 1.27), but low expertise level (Mean=9.96, SD 6.74), which also indicated the lack of necessary in-service training opportunities.

Usage of ICT in education is a complex process where many agents play different roles. Forces that may influence or impede ICT usage outside formal schooling should be taken into consideration. Contrary to this fact, much of the early research on computer use in education has ignored teachers’ attitudes toward the new technologies. Studies focused on ICT and their effect on teacher’s competence; thus overlooking the psychological and contextual factors involved in ICT applications. However, it should not be forgotten that successful implementation of educational technologies depends largely on the attitudes of educators, who ultimately determine how they are used in the classroom: teachers’ attitudes are the major determining factor in the adoption of technology. According to Rogers (1995), people’s attitudes toward a new technology are the key element in its diffusion. Since Rogers uses the terms innovation and technology interchangeably (p. 12), the diffusion of an innovation framework seems particularly suited for the study of the diffusion of ICT. This suggests that studies at the early stages of technology implementation should focus on the end-users’ attitudes toward technology.

A significant relationship has been found between the proximity of computers and the number of access resources (both at home and school) on the one hand, and, on the other, teachers’ attitudes toward computers. Teachers who have a high perception of efficacy tend to use computer related tools in classroom more frequently than the others. The results showed that teachers who have computer at home have a higher level of expertise than the others.

All the results reported brought us to a conclusion which is also stated in the literature by many researchers: the goal of the integration of technology into the social studies area, like in all other areas, has yet to be reached (Barron, Kemker, Harmes & Kalaydjian (2003); Bielefeldt (2001); Mills & Tincher (2003)).

Conclusion

By the 1930s, Turkish schools had teaching materials such as maps, laboratory equipment, and filmstrip projectors for instructional use. Although they had teaching material such as maps and other equipment they did not use them. Until the 1940s, mostly printed instructional materials were used in schools. Between 1950 and 1970, schools had technologies such as audio cassettes and overhead projectors. During the 1970s, several new teaching materials were provided for schools and introduced to teachers. In addition, some big universities started to offer graduate programs aimed at training professionals in the field of traditional educational technology. Though some of these traditional technologies are still in use to prepare students, educational policy makers in Turkey believe that schools must give
students the knowledge and skills they will need in the future. Therefore, computers have gained more importance than any other educational technology (Usun, 2004).

In order to improve ICT facilities and the skills of teachers, the Ministry of Education is making investments, providing in-service training and providing accessibility to resources. Besides equipping selected schools with IT classrooms, another arrangement was made to provide accessibility to these resources to other schools nearby. Furthermore, the Ministry of Education has planned a project to provide each teacher with his/her own laptop computer. These kinds of innovations will support all teachers in Turkey in becoming competent teachers in terms of ICT usage, besides furnishing access to the many other advantages that technology provides. The diffusion of many new technologies in society has not been equitable. Rogers (1995) theorized that people who are innovative and quick to adopt new technology tend to be younger and better educated, and to earn higher incomes than later- and non-adopters of technology. Socio-economic status, access to resources and equity in outcomes are important issues in education. For instance, the availability of computers at school enables many teachers to use them even though they may not have a computer at home.

Developing countries are vitally dependent on substantial foreign assistance to ensure the development of ICT. However, it is often very difficult to persuade donors to focus on ICT. These countries are perennially short of foreign exchange to acquire the latest technologies. Most developing countries are undergoing ‘Structural Adjustment Programs’ under the auspices of the IMF. Cost-efficiency of an ICT is another major factor that is important in determining its growth. Developing countries have to ensure that the technology that is adopted is easily accessible to the target group and also fulfills all the functions that are expected of it. Such a scenario essentially implies that a costly technology need not always be the best technology. However, it is often seen that developing countries invest in the latest technologies without considering whether the target audience can be reached effectively or whether the target audience is interested in the technology (Usun, 2004).

In the last decade, there have been a number of parallel projects related to the integration of ICT in Turkey’s educational system. At the end of National Education Development Project (with the World Bank and HEC), faculties of education reconstructed their curriculum to train teacher candidates with abilities and skills to use ICT effectively in their subject areas in 1998. After this year, the number of ICT projects increased. For instance, the first phase of the basic education project (with the World Bank) started in 1998. The scope of the project was to build information technology classrooms in at least two primary education schools in 80 cities and every town, and the identified schools were grouped according to the number of students. In that context, 2,834 information technology classrooms have been scheduled to be built in 2,451 primary education schools all over the country. This number has increased to 2,802 with 351 newly constructed schools. The establishment of information technology classrooms in these schools has been completed in all cities and towns (Goktas & Yıldırım, 2003).

The key points of this study on the ICT usage of social studies’ teachers are as follows: As a developing country, Turkey is attempting to foster a culture of acceptance amongst the end-users of ICT. Therefore, teachers’ attitudes are indispensable to ICT usage in the classroom. The key is double-headed: on one side there is the human factor and on the other is the technological factor. Some limitations such as teachers’ lack of ICT usage skills and insufficient infrastructure supporting ICT impede ICT usage. Hence, it is necessary to promote these skills as a prerequisite to delivering ICT facilities. The other main barriers to the implementation of ICT as perceived by the teachers in this study are the mismatch between ICT and the existing curricula and the class-time frame. It follows that placing ICT in schools is not enough to attain educational change. The introduction of ICT into education requires equal level of innovation in other aspects of education. Both policymakers and teachers share this responsibility. Policy-makers should provide additional planning time for teachers to experiment with new ICT-based approaches. Reducing the teaching load for the teachers may attain this. Teachers’ preparation necessitates not merely providing additional training opportunities, but also aiding them in experimenting with ICT before being able to use it in their classrooms. If decision makers want to involve teachers in the process of technology integration, they need to find ways to overcome the barriers perceived by the teachers.

Suggestions

Here are some suggestions for effective usage and implementation of ICT. It would be useful to provide ICT knowledge as modules so that teachers may integrate ICT into lessons. The ICT materials should therefore be based
on classroom research and provide excellent teaching ideas and activities for developing and strengthening students' concepts, skills and meta-cognition. Teachers and researchers can make even better use of ICT facilities together. ICT has vast potential in education but its effective use must be carefully tried out and planned by researchers and teachers who know what to do with it in the teaching-learning process. They have to determine what strategies are needed for certain learning situations and how learning processes can be enhanced using the technology. Merely providing schools with hardware, software and in-service training is not enough. Any in-service training needs follow-up support, peer coaching and peer dialogue to ensure successful use of the new technologies. Teachers must be part of the decision making process with respect to the implementations of ICT innovations in schools, so that they may commit to the innovation with conviction. Easy access to databases of the available curricular resources and strategies would provide very useful information for communication among teachers. There is a need for localized resource centers to provide support for schools within a certain district. Workshops for teachers and students in the area could be conducted to promote cooperative projects and sharing of experiences and expertise.

Most teachers rely heavily on textbooks and blackboards; we can re-vitalize education with ICT. In subsequent training workshops, key personnel and trainers could each have opportunities to present an aspect of ICT concept or use in the classroom. These presentations could be about a feature of some ICT concept, resource or application in the teaching and learning of a subject, a useful web site, or to show the work of pupils themselves. The presenters could bring along handouts for every teacher as well, so that teachers can compile the materials into a guide or resource book. Repeat sessions may be necessary if there are too many teachers for one group. There are also difficulties with the maintenance of hardware and the purchase of new equipment and software because of the high costs involved. It would also be useful to have technicians available to assist the schools. Educators find it difficult to integrate and introduce ICT for everyone in schools due to the high cost of the provision and updating of networked equipment needed in an already overloaded curriculum, and with teachers knowledgeable in ICT in short supply. Teachers stated that they attended some courses, which the principals selected for them. The instructors were not competent in using computers and their knowledge of computer was average. Providing continuing courses and sharing workshops planned throughout the implementation of the project might be useful for these teachers. Independent teachers will be able to achieve the project objectives with minimum support from key personnel. A regular newsletter or bulletin would be useful for teachers to write about good teaching ideas, that they have found to be effective and successful, and for sharing with others throughout the country and beyond.

References


Watson, D. M. (1998). Blame the technocentric artifact! What research tells us about problems inhibiting teacher use of IT. In G. Marshall, & M. Ruohonen (Eds.), *Capacity Building for IT in Education in Developing Countries* (pp. 185-192), London: Chapman & Hall.


Appendix (Information and Communication Technology Usage Survey)

(This survey is presented here without directions and “other please specify” items appearing under each heading as the last item)

Do you have your own computer?  Yes ( )  No ( )
Do you have computer at school?  Yes ( )  No ( )
Do you have Internet connection at home?  Yes ( )  No ( )
Do you have Internet connection at school?  Yes ( )  No ( )
Daily Computer Usage:  Less than one hour ( )  1-3 hours ( )  3-5 hours ( )  More than 5 hours ( )

<table>
<thead>
<tr>
<th>Software Usage</th>
<th>Good</th>
<th>Average</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Processors (Word etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spreadsheets (Excel etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation Software (PowerPoint etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Databases (Access etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Aided Instruction Software</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web Page Development Tools (FrontPage, dreamweaver etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web Browsers (Netscape, Explorer etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Engines (google, yahoo etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic Mail (e-mail)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion Lists and Newsgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chat and/or Forum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic Encyclopedia and/or Atlas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional Films (video, CD, VCD etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Usage of Instructional Tools and Materials</th>
<th>Frequently</th>
<th>Sometimes</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Board</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overhead Projector</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opaque Projector and/or Document Camera</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multimedia Computer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer – Projector System</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet/Web Environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Television/Video</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radio Cassette Recorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Camera</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slide Projector</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Printed Materials (journals, books, worksheets etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Professional Development about ICT</th>
<th>I Prefer</th>
<th>Neutral</th>
<th>I don’t prefer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Printed Materials (manual or journal etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self experiment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participating seminars or taking courses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Resources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support Resources</td>
<td>I Prefer</td>
<td>Neutral</td>
<td>I don’t prefer</td>
</tr>
<tr>
<td>Experienced teachers on ICT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colleagues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other colleagues in different schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical support units in the schools</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factors Encourage Technology Usage</th>
<th>Important</th>
<th>Neutral</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rewarding the technology usage efforts of teachers in instructional activities</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

49
Investments of the institution on infrastructure of instructional technologies
Investments of the institution on in-service education programs for instructional technologies
Investments of the institution on the support services of instructional technologies
Developing the policies and plans for diffusion of the instructional technologies
Providing support for the projects towards the expansion of instructional materials
Carrying out the studies for integration of technology into curriculum
Reducing work load to provide opportunities to teachers for developing instructional materials

<table>
<thead>
<tr>
<th>Perceptions about use of ICT</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t use computers as much as other resources (books, overhead projectors etc.) for instructional purposes.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I know what to do for using computers in instructional environments.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am aware of the opportunities that computers offer.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can answer any question my students ask about computers.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am not sure that I am computer-literate for use computers in my classes.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don’t want to use computers.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that I can use instructional technologies in class activities more effectively day by day.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe that tools like e-mail, forum and chat will make communication with my colleagues and students easier.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that technology supported teaching makes learning more effective.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think the use of instructional technologies increases the interest of students toward courses.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think the use of instructional technologies increases the quality of courses.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that usage of instructional technologies makes it easier to prepare course materials (assignments, handouts etc.).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is hard for me to explain the use of computer applications to my students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can handle different learning preferences of my students having different learning styles by using instructional technologies.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think technology makes effective use of class time.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think using instructional technologies makes me more productive as a teacher.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that using technology makes it easier to reach instructional resources.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don’t prefer to be assessed about my instructional technology based applications by any other professionals.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Barriers to Technology Usage</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inefficient time to prepare materials based on technology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inefficiency of teachers’ technical knowledge to prepare materials based on technology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problems about accessibility to existing hardware (computer, overhead projector etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inefficiency of institutions computer laboratory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inefficiency of institutions technical infrastructure about instructional technology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inefficient number of media (printer, scanner etc.) for effective use of computers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortage of computers used by teachers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence of reward systems for encouraging technology usage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor technical and physical infrastructure of learning environments.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inadequacy of computers used by learners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inefficiency of guidance and support by administration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insufficiency of financial resources for technology integration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inefficiency of instructional software/electronic resources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scarcity in resources on technology for attaining information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficiency in professional development opportunities for gaining knowledge and skill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficiency in support services in material development/technology usage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of interest of teachers in technology usage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulties of improper teaching methods for technology usage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inadequacy of the courses of technology offered to students</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Study of Uses of ICT in Primary Education through Four Winning School Cases in the Taiwan Schools Cyberfair

Shelley S.-C. Young and Hsin-Ho Ku
Graduate Institute of Information Systems and Applications & Centre for General Education, National Tsing Hua University, Taiwan // scy@mx.nthu.edu.tw // juby@cht.com.tw

ABSTRACT

The purpose of this study was to understand, describe and interpret the uses of information communication technology (ICT) in primary education in the Taiwan Schools Cyberfair as a means to expand and enhance student learning in an extra-curricular setting. Four winning schools, with 48 teachers and students involved, were purposefully selected for study. The research findings revealed that students in each group were highly motivated through the extensive use of ICT. Through long-term engagement with ICT, students were empowered to conduct both wider and deeper exploration of their selected topics of interest. We report and describe the effects of using ICT. In this study, three collaborative learning models for applying ICT in project-based learning (PBL) mode are identified in the specific contexts: the SCGP model (same class, grade, and project); the DCGSP model (different class and grade, same project) and the DCGP model (different class, grade, and project). In addition, we identify a new and important collaborative model, SPECS, which stands for school (intra), parents, enterprises, community, and school (inter). The SPECS model connects all the possible participants involved in the ICT-enabled activities that contributed to the success of the outstanding performances of the four winning schools at the Schools Cyberfair contest. In the conclusion we discuss some specific issues relating to PBL.

Keywords
Technology-facilitated learning in complex domains, Project-based learning (PBL), Web system supported learning, Online learning, Learning effects, Collaborative learning

Introduction

With the advent and pervasiveness of information communication technology (ICT), the Internet has been widely accepted as an important tool in life and education in recent decades. Active diffusion of information technology through investments in national information infrastructures and education has become a world trend (Halverson & Collins, 2006). In order to adapt to this technological innovation in education and to keep abreast of world trends, Taiwan’s Ministry of Education (MOE, 1999; 2003) holds the diffusion of IT in education as one of its major tasks and has invested a great deal of money towards it since the early 1990s. Consequently, the availability of computer equipment is not a major problem in facilitating IT in education in Taiwan. As a result, the facilities in most of the computer labs and for teacher training have been greatly improved. It is anticipated that education reform can be further accelerated and learning opportunities increased.

However, in Taiwan the traditional goal of education and the prevailing social value of obtaining a high grade continue to dominate the core practices of schooling (Young, 1999a, 1999b, 2000a, 2000b, 2001, & 2006). Observations of school practices indicate that the national policies on implementing IT in education, such as the inclusion of supplementary digitized learning resources, have indeed had some impact on teachers and motivated them to change. In general, however, there is not much widespread use of technology practices in most schools’ regular curricula.

Overwhelmingly, core instructional practice is still examination-oriented, and the use of ICT in teaching practice in schools remains marginal. Most of the time, the positive effects of applying IT arise from specially funded experiments during certain periods of time. Those teachers who were early adaptors of technology were more likely to adapt to the change by taking on projects such as providing additional alternative options for students (Young, 2006).

ICT in education is being promoted and studied in various countries and, in general, these countries’ goals are similar: to provide more effective learning and competitive manpower in the international market. However, research conducted in any one country may not yield results completely applicable to other countries because of cultural
differences, educational traditions, economic status, or political priorities. Every country has its contextually specific problems (Belland, 1998). This study reports on an investigation of uses of ICT in primary education in the Taiwan Schools Cyberfair as a means to expand and enhance student learning in an extracurricular setting. It documents how ICT has been integrated into an alternative instructional setting through project-based learning (PBL) and cooperative learning. It puts forward the case that, ultimately, ICT enables learners to open up a new window to the outside world and bridge the information divide.

**Literature review**

Project-based learning (PBL) is a constructivist pedagogy and class-oriented learning approach. Unlike traditional learning which is short-term, subject independent, teacher-focused, and mostly constrained to classroom settings, PBL involves long-term, theme-based learning and student-centred activities that focus on daily life problems and allows learners to use an inquiry-based approach to engage with issues and questions that are real and relevant to their lives. When conducting PBL, teachers encourage students to choose topics of their own interest, and set specific questions in a well-planned framework that is wider than the immediate task at hand. Students can gain knowledge, ask questions, and find solutions during the research process (Curtis, 2001; GLEF, 2001).

In project-based learning, teachers actively supervise students by taking the role of project facilitator instead of instructing them (Thomas, 2000; Curtis, 2001; Wong et al., 2006). Teachers support students not only as a source of knowledge, but also as co-learners/peers in their activities. In a PBL environment, teachers are no longer the centre of learning. Rather, students design their own activities and answer driving questions. Through the process of question-raising, cooperation, data collection, communication, and result demonstration, a highly engaging atmosphere and rich learning environment that focuses on students is thus created (Marx et al., 1997). Moreover, teachers in a PBL environment are not only leaders in the classroom but also do a more complex and important job. They are not only course designers or assistants in learning activities but also judges to evaluate learning effects (Delisle, 1997). In PBL, by asking questions, cooperating with others, analyzing data, and communicating with each other, students can create a student-centred learning environment or learning community (Blumenfeld et al., 1991; Marx et al., 1997; Erstad, 2002).

The long-term studies conducted by the George Lucas Educational Foundation (GLEF, 2001) conclude that PBL has many benefits for students, including:

- deeper knowledge of subject matter;
- increased self-direction and motivation;
- improved research and problem-solving skills.

According to the study of Challenge 2000 Multimedia Project by SRI (2000), the research findings indicate that student participation in learning activities changed through their involvement with PBL. Teachers reported that their students became more self-regulated learners, engaging more actively in classroom learning, taking responsibility for their learning, and becoming more skilled collaborators with their peers (SRI, 2000). In other words, the students’ PBL activities displayed increased motivation, increased responsibility for their own learning, better peer collaboration, improved content mastery, better understanding of target audience, greater self-confidence and self-esteem, more peer teaching, better technology skills, more time on task, and more skill in analyzing and problem solving (SRI, 2000; Ullah, 2003).

In the information era, when ICT is used in a meaningful way in project-based learning, students work in teams to conduct research using a variety of sources, ranging from digital information on the Internet to interviews with selected interviewees. They explore real-world problems over an extended period of time and ultimately create digitized presentations to share what they have learned (Curtis, 2001; GLEF, 2001; Wong et al., 2006). Students’ learning benefits include increased confidence in both written and spoken communication and ICT skills, as well as enhanced project skills. Students learn how to build, manage, and share their web-based resources within and across the project groups.

Johnson and Johnson (1994) point out that collaborative learning allows learners to work together to achieve mutual learning goals. Sharan and Shaulov (1990) suggest that collaborative learning increases interdependency and helps fulfil learning goals through cooperation and assistance among learners. Students can acquire the abilities of
communication and coordination. They are willing to comment on each other’s work, share personal viewpoints, express their own opinions, and learn to accept views different from their own. Through organized and systematic teaching strategies, teachers are able to group students of different abilities, sexes, and backgrounds together to learn jointly, share experiences, and receive recognition from peers (Johnson & Johnson, 1994). Individual learning effects are upgraded by interesting activities and interaction among group members.

In sum, we know that collaborative learning is a well-arranged and methodical teaching strategy that improves student learning. In this cooperative model, students help each other in the learning activities. Also, students’ self-esteem is enhanced in their academic establishment. Meanwhile, students learn how to respect others (Slavin, 1985). Therefore, with various learning activities, effective learning is obtained by improving students’ motivation, academic performance, and problem-solving skills.

Project-based collaborative learning uses ICT to arouse students’ motivation to learn, impelling students to participate actively in online discussions and deep research (Polman & Fishman, 1995; Erstad, 2002). They collaboratively investigate questions of daily life and work with peers or teachers for the solutions (Blumenfeld et al., 1991).

Although the Taiwan Schools Cyberfair events draw much participation and attention nationwide, not much related research work has been conducted on this basis. Therefore, through in-depth study of the four representative elementary schools that have performed outstandingly in the contest, we hope to shed light on the uses of ICT with PBL and try to identify the collaborative models employed in the extra-curricular educational settings supported by ICT.

**Background to this study: About the Taiwan Schools Cyberfair**

The Taiwan School Cyberfair is based on the International School CyberFair, a global contest in designing project-based learning (PBL) websites put on every year since 1995 by the Global SchoolNet Foundation. Winners selected by both peer evaluation and experts are awarded certificates, money, and prizes. Their achievements are displayed on the official website, available to other students all over the world. Since 2000, the contest in Taiwan has been held once a year to promote information education, reduce the digital divide, encourage new teaching practices, facilitate cultural exchange, and broaden students’ international awareness through the use of ICT in their learning and living settings. The official website provides students with system support helpful to their projects. Students wishing to join the contest are required to register online and, after registration, they would be validated to become formal participants of the contest. The interactive modes of PBL supported by ICT are shown below in Figure 1.
The extensive application of the PBL approach in Taiwan can be identified among the activities of the Taiwan Schools Cyberfair. The Cyberfair contest, based on PBL, has both global and local significance. Students participating in the Schools Cyberfair work in small groups. They collaborate with each other, facilitated by teachers. Each group is assigned a task, and the members cooperate to fulfil it. Along the way, through the process of repetitive discussions, students achieve more effective learning results than they would have through individual learning. This type of project provides students with the opportunity to obtain a more in-depth understanding of their community through project-based action studies and to present outcomes via multimedia on the Internet. The activities incorporate valuable cooperative learning within and between countries and communities. Ultimately, the localized content pertaining to different countries displayed on the Internet helps facilitate the development of global concerns and visions.

**Purpose of the study**

The purpose of this study was to investigate the alternative ways of applying ICT in primary schools in informal educational settings in Taiwan. The results are anticipated to shed light on the potential applications of ICT in the Schools Cyberfair, which offers the students a stage to present their technological competence and collective web projects to their peers as well as to seek a flexible combined learning approach and model through the uses of ICT.

**Research questions**

In order to explore the potential uses of ICT in primary education for alternative learning opportunities, the study asked the following specific questions:

1. Since participating in the Schools Cyberfair contest requires group commitment and long-term (at least six months) engagement, what are the possible student motivations in being actively involved in the web-based project contest?
2. What might be the teachers’ cooperative model of adopting PBL? What might be the collaborative models applied by the winning teams in producing web projects with the support of information technology? What might be the overall cooperative model identified from the contest?
3. What are the students’ learning gains in this kind of extra-curricular learning setting?

**Methodology**

The methodology employed in this study is primarily a qualitative approach. This study is interpretive and descriptive in nature and uses the case-study method. Case studies are used to examine a specific unit such as an event, a programme, an organization, and a time period in depth and detail, in context, and holistically (Merriam, 1988; Patton, 1980; Stake, 1995). In addition, quantitative data, such as student demographic information, is also included. The researchers conducted individual and group interviews with the students and the teachers at school sites. Online questionnaires requesting personal and background details of the teachers and students (demographical data such as sex, age, computer skills, and Internet use) were administered to the participants at the beginning of the study. In addition, observations were conducted intensively on the Internet, focusing on subjects’ online PBL learning activities, including the collection of their electronic data from journals, chats, discussions, and email exchanges on the Internet. The research was also concerned with how the students collectively designed and presented their web materials and how they interacted with the distributed students. Triangulation was used to improve the probability that findings and interpretations would be reliable.

**Data collection**

The data-collection methods and tools for this study included a questionnaire, telephone interviews, online data collection, activity observations, and quantitative data analysis of the web projects produced by each group.


**Duration of the study**

The study lasted about a year and a half, from September 2003 to February 2005 and comprised three major phases:

1. Preparation and orientation: September 2003 to January 2004: participants’ school learning activities in preparation for the Cyberfair contest;
2. Participation and collaboration: February 2004 to July 2004: Participants’ collective effort in the web project, online journal keeping; and
3. Follow-up data collection: August 2004 to February 2005: follow up interviews with the students and teachers at school sites.

**Data analysis**

The data analysis procedure included two main phases: the descriptive statistical data analysis and the qualitative data analysis. Regarding the descriptive data, SPSS software was used for data storage, and for the calculation of frequencies and percentages. The qualitative data analysis was guided by the research questions stated previously. The analysis evolved around the data reduction, organization, and matching as well as the generation of categories that resulted from the study of all data sources. Questionnaires and interviews were the primary data source for this study. Observation notes, online journals, related documents, and web logs were used to provide an extensive understanding of the study and PBL contexts. Demographic data related to each case school’s background information and information technology status and students’ profiles were reviewed. The categorized data eventually assisted the researchers in data interpretation, drawing conclusions, and verification.

**Limitations of the study**

The present study has certain limitations that need to be taken into account. The four winning cases were carefully selected for study. The number is too small to be representative of the elementary-school population in Taiwan so the findings cannot be directly generalized to the larger population. The validity and the reliability of the study are limited by the level of honesty in the participants’ responses to the instruments.

**Results and discussion**

For a better understanding of the four school cases and participants, we first focus on the participants’ profiles and then address the research questions, including looking at students’ motivations in participating in the PBL web-project contest, the SPECS collaborative model derived from the PBL supported with ICT, three PBL-collaborative models enabled by ICT, learning effects of PBL activities, and analysis of applications of ICT in project-based learning.

**Participants and four school cases**

Four winning schools from the Taiwan Schools Cyberfair contest were identified as four cases for in-depth study. They were selected according to their ethnic origins, regions, contest themes, and willingness to participate and came from different parts of Taiwan, representing eastern, southern, rural, and urban areas. The four winning schools together comprised 48 subjects (36 students and 12 teachers). They participated in the theme of “local features for sightseeing.” Their schools and locations are listed in Table 1. To preserve subjects’ privacy and confidentiality, the information has been coded.

Each of the groups consisted of eight to ten students, with three teachers as advisors (Table 1). The teachers in this study consisted of nine males and three females. Their age range was 25 to 46 years, with a mean age of 29.58. They created a PBL learning environment, helping students to create their own web project.
Table 1. Basic information for the four schools and participants

<table>
<thead>
<tr>
<th>School Code</th>
<th>Location</th>
<th>Total number of students in each school</th>
<th>Students participating in PBL projects</th>
<th>Teachers participating in PBL projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>School B</td>
<td>Northern mountain area</td>
<td>65</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>School H</td>
<td>Northern city area</td>
<td>1985</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>School U</td>
<td>Central rural area</td>
<td>1800</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>School K</td>
<td>Off-shore remote island</td>
<td>134</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total = 48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Of the four schools, B, H, K, and U, two (School B and School K) were mini schools with 65 to 134 students. School B is in a remote mountainous area in northern Taiwan. Most of its students are aboriginal. School H is located in the town of Dan-shui, in Taipei County. School U is in the village of Yong-jing, in Chaung-hua County, which is in the central part of Taiwan. School K is in Jin-men County, on an off-shore remote island.

Information technology status of each school and the students

The information technology status of each school and the subjects will help us understand more about the computer access of those students in this study. Table 2 below indicates the school locations, computer facilities, and percentage of participants’ with personal home computers.

Table 2. School location, computer classrooms and the rate of ownership of personal computers

<table>
<thead>
<tr>
<th>School code</th>
<th>Number of computer classrooms/computers</th>
<th>Number of participants’ personal home computer ownership/rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/18</td>
<td>3/10 (30%)</td>
</tr>
<tr>
<td>2</td>
<td>2/80</td>
<td>6/8 (75%)</td>
</tr>
<tr>
<td>3</td>
<td>1/38</td>
<td>8/10 (80%)</td>
</tr>
<tr>
<td>4</td>
<td>1/32</td>
<td>8/8 (100%)</td>
</tr>
</tbody>
</table>

Among the three schools in Taiwan (Schools B, H, & U), the percentage of the participants’ personal home computer ownership in School B (30%) is far less than that of School H (75%) and School U (80%). It seems that the IT resources vary according to school locations. However, the technological equipment in school K, on the off-shore island, is almost the same as that in the other schools in Taiwan. Surprisingly, the participants’ personal home computer ownership in school K is 100%, even higher than the percentage of ownership among students from the schools in Taiwan. This indicates that the parents in school K show great awareness of the importance of information technology in their children’s lives. This point was further verified by interviews with teachers who reported that the parents in school K offered great support and assistance to the students during the project-making process. Conversely, parents at the other schools tended to be less involved in students’ project-making because they were too busy with work.

Student motivations for participating in the PBL web project contest

According to the data collected from the student questionnaires, interviews, and observations, student motivation to be engaged long-term in the PBL web project contest held by the Cyberfair can be categorised into five types:

1. **Personal interest.** According to our analysis of the questionnaires, almost all of the students found the PBL scenario combining teaching and outdoor activities in the real world to be very “interesting” and “fun” (B04, B05, B06, B07, H01, H04, H06, H08, U04, U05, U09, U10, K02, & K05). Students U01 and H03 were also curious about a method of doing project research that was not common in regular lecture-based teaching settings. Their
learning interests were piqued and they were excited to take part in related learning activities outside of school, during their leisure time. Data from the questionnaires revealed that the students gave overwhelmingly positive feedback to PBL, combining the networked system offered by the contest organizer and noticing that they had gained a lot through the PBL activities. The following comment reflects this sentiment: “I found that this activity is great, even if I didn’t win the prize… Learning how to make use of the Internet correctly and knowing more about computer technology made our life rich and interesting!” (H05).

2. **Gaining situated knowledge and skills.** Students got involved in authentic affairs in the real world and outdoors through PBL activities. In conducting a PBL web project, they had to collectively make a plan in advance, allocate individual work, find supporting resources, go out to collect documents, take pictures, and interview people using digital cameras and recorders, etc. They applied their newly acquired IT skills and knowledge in real-life situations. The students felt that they learned new knowledge and skills from the extracurricular activities (B09, B10, H01, H04, H05, H06, U02, U03, U04, U05, U07, K01, & K06) (Brown, Collins, & Duguid, 1989).

3. **Winning awards.** The main motivation for most of the students to participate in this project-making activity was to win a prize. Some students said that if they were to win, they would feel very proud because they would “receive the awards and medals in public” (U05 & H05). The goal of achieving recognition appeals to the students’ inner sense of honour, impelling students to participate actively in the contest (U04, U05, H05, K05, & K07).

4. **Encouragement from teachers or schools.** There are two kinds of encouragement: teachers’ encouragement to students; and the encouragement from school administrative officials.
   1) Teachers’ encouragement. Teachers in the four schools actively asked their students to take part in the contest and pushed them to make the projects with additional support during out-of-class time.
   2) Encouragement from the school. For example, the principal of School B has encouraged students to participate in the contest since 2000. At Schools H, U, and K, the school officials offered a lot of assistance in obtaining software and hardware as well as providing other resources for the students.

5. **Helping boost local tourism via the Internet communication.** The four groups in this study all signed up in the “local tourist resources” category of the Taiwan Schools Cyberfair. The purpose of producing their web projects was to present the natural beauty and sight-seeing spots in their hometown. In order to give an overview and provide in-depth information about their chosen sites from different aspects, students in each group had to study historical events, culture, the origins and significance of local spots, and agricultural projects in order to help boost local tourism through the Internet, free of charge. In our follow-up interviews with the students and the teachers, they mentioned that the project raised people’s awareness of the community and their identity within their neighbourhood (B05, B09, H03, H05, U03, U05, U07, U09, K01, K02, K07, & K08). They hope to advance economic development and increase local prosperity through this IT-boosted event led by the younger generation.

**Teachers’ collaborative model of structuring PBL with the students**

According to our interviews and observations, we have identified three major roles that the teachers played during the projects:

1. **Project leader.** The project leader is in charge of project integration, planning and, knowing students’ abilities, working with students to decide the project direction and goals and set the timeline of the project. This role was dominant and important to the success of the project.
2. **Website technician.** The website technician’s role was to mainly give professional opinions on making the website, instruct students on how to use all kinds of editing software and hardware, maintain computer facilities, and solve problems related to website creation.
3. **Coordinator.** The coordinator was in charge of external and internal communication, combining ideas of students and other teachers and communicating with related organizations.

The three kinds of teachers had their own contributions to the projects. They worked together to assist students make the project run smoothly (see Figure 2).
Three PBL collaborative models enabled by ICT: SCGP, DCGSP, & DCGP

By analyzing the make-up of each group, the observed data, and student ways of collaboration, we identified three collaborative models: (1) SCGP Model (same class, grade, and project), (2) DCGSP Model (different class and grade, same project), and (3) DCGP Model (different class, grade, and project). (See Table 3.)

<table>
<thead>
<tr>
<th>Models</th>
<th>Ways of collaboration</th>
<th>Representing schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCGP</td>
<td>Collaboration with classmates (same class, grade, and project)</td>
<td>Schools U &amp; K</td>
</tr>
<tr>
<td>DCGSP</td>
<td>Collaboration with different grades (different class and grade, same project)</td>
<td>School B</td>
</tr>
<tr>
<td>DCGP</td>
<td>Collaboration with different grades and projects (different class, grade, and project)</td>
<td>School H</td>
</tr>
</tbody>
</table>

During the PBL process, it came to our attention that the students in the four schools had the flexibility to break the class or grade constraints by expanding their cooperation beyond their own classes. Use of the networking technology meant that students could form a project group transcending the limitations of classes, grades, space, and
time — something that was otherwise not possible before the advent of ICT. Excitedly, they communicated, shared information, and exchanged digital files on the Internet. These kinds of cross-class or cross-grade cooperative groups gave students more stimulus and experiences, allowing students to create different learning environments in the given situations.

1. **SCGP model (same class, grade, and project)**

The groups in schools U and K belonged to the same class. Three teachers helped organize group members to work on the project together after school. The members communicated and shared individual results through e-mail and networked file-transferring software. For those teachers, the Internet became an important tool to monitor student progress and give students advice on time management. This model is common nowadays in school settings, which makes it easier for teachers to teach and easier for students to interact with each other (Figure 3).

2. **DCGSP model (different class and grade, same project)**

The working model of school B showed that the leading advisor was a fifth-grade teacher. After discussing with students, the advisor led the group to initiate their project and conduct interviews. A sixth-grade teacher was involved in this project, too. Three sixth-grade students were invited to work with them. Therefore, the project was made by a group of fifth-grade and six-grade students who made concessions in their leisure time to work on the project. In this case, the students were tied closely to the project, having incredibly great interaction (Figure 4).

![Figure 4. DCGSP model](image)

3. **DCGP model (different class, grade, and project)**

The collaborative DCGP model (Figure 5) of school H was a cross-grade integration. Because of the great promotion of the competition by the school, there were up to four groups participating in the web project contest. The group in our study was composed of fifth-grade students, while the other three groups were made up of sixth graders. Though the project title of each group was different, the content was centred on the natural and cultural neighbourhood of the school, which resulted in repeating the same interview subjects. Thus the data collected by different groups could be integrated and shared. Juniors and senior students could exchange resources, share information, and learn techniques by e-mail and FTPs. The model made the process more competitive, and enabled the students to be actively involved in their project (HT1).
The researchers noticed that the project participants of schools B and H went through all kinds of cross-grade learning activities. Both teachers and students had to deal with the difficulties of time, space, student proficiency, and age difference. However, the PBL learning environment that combined the applications of networked technology made this possible. Furthermore, in addition to the teachers, the senior students also helped the younger ones with their learning. The students could enhance their ability by stimulating or challenging each other. The new cooperative models shed light on different approaches of learning and teaching with the integration of ICT in traditional educational settings.

The SPECS collaborative model of PBL supported by ICT

The data collected from our prolonged engagement and persistent observations in this study supported the SPECS collaborative model of PBL supported by ICT. The SPECS collaborative model, standing for school (intra), parents, enterprises, community and school (inter), connects all of the possible participants involved in the ICT-enabled activities that contributed to the success of the outstanding performances of the four groups at the contest. The SPECS collaborative model is shown in Figure 6.
We noted that when the students and teachers were carrying out PBL activities on the web projects, there was a lot of assistance from teachers, schools, parents, communities, and other groups. The phenomena indeed are not common in most of the communities or our society in Taiwan. However, the online contest, visible and accessible to most people, has changed the atmosphere and has also piqued people’s interests and involvement regarding the activities. The community and parents worked together in securing travel arrangements and safety for the students’ field trips. Local literature and history resource centres cooperatively provided students with precious historical data and pictures. For example, when students of school B encountered linguistic difficulties in communicating with tribal elders in their mother tongue, parents volunteered for interpretation and data collection. Schools U and H obtained a lot of important data from local literature and history workshops, local enterprises, and civil, cultural, and educational institutions. Impressively, the members of school K on Jin-men Island received help not only from local people, but also from residents from across the strait in Taiwan.

With PBL, students could expand their learning arena to places off campus and in the community. The assistance from the schools, parents, community, and other groups enabled the students not only to seek resources and help from outside of the school campus but also to gain communication skills through the process of talking to elders or strangers.

Learning effects of PBL activities and network system to support PBL activities

Both the statistical and qualitative results of the learning effects of the PBL web project and the effects of using the information system to support PBL activities are presented below. These results are based on the questionnaires, students’ online journals, and interviews with the students.

The questionnaire was composed of nine 5-point Likert-scale questions, in two parts: students’ perceptions of the learning effects of PBL activities and students’ perceptions of the effects of using the information system to support PBL activities. A total of 30 questionnaires out of 36 were collected and used for analysis. The descriptive statistical results of this questionnaire are reported below (Table 4).

<table>
<thead>
<tr>
<th>I. Students’ perceptions of the learning effects of PBL activities (N = 36) (%)</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I could learn in-depth information systematically through group work by means of the PBL approach.</td>
<td>33.3</td>
<td>53.3</td>
<td>13.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2. I have experienced new information presentation skills collaboratively in the process of doing the PBL project.</td>
<td>63.3</td>
<td>26.7</td>
<td>10.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3. I could apply computer skills and literacy gained in the class to the PBL web production project in the extra curriculum setting.</td>
<td>33.3</td>
<td>46.7</td>
<td>20.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4. I acquired much more knowledge and applied new writing skills beyond the regular classes in the process of doing PBL project.</td>
<td>60.0</td>
<td>30.0</td>
<td>10.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5. I have learned and applied a lot of communicative skills in the process of doing the PBL project.</td>
<td>30.0</td>
<td>50.0</td>
<td>20.0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Students’ perceptions of effects of using the network system to support PBL activities</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I have learned to apply a lot of information communication technology (ICT) skills through engaging in the PBL project in designing and producing the webpage.</td>
<td>46.7</td>
<td>40.0</td>
<td>13.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7. I have used the information system on the Internet frequently in the process of doing PBL project and thus my WWW information and my personal</td>
<td>46.7</td>
<td>40.0</td>
<td>13.3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
knowledge has been increased and enhanced.

8. The use of the information system on the Internet helped me improve my learning effectiveness in doing PBL project.

9. Reviewing the other groups’ web projects displayed on the “works exchange area” and “exhibition area” on the Internet as a result of the PBL contest has increased my personal knowledge and ICT skills and upgraded my views.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8.</td>
<td>The use of the information system on the Internet helped me improve my learning effectiveness in doing PBL project.</td>
<td>43.3</td>
<td>40.0</td>
<td>16.7</td>
</tr>
<tr>
<td>9.</td>
<td>Reviewing the other groups’ web projects displayed on the “works exchange area” and “exhibition area” on the Internet as a result of the PBL contest has increased my personal knowledge and ICT skills and upgraded my views.</td>
<td>50.0</td>
<td>36.7</td>
<td>13.3</td>
</tr>
</tbody>
</table>

According to the students’ self-reported data collected by the questionnaire, the nine statements listed in the table received overwhelmingly positive responses from the students. The data indicate that students could benefit from the PBL approach in doing the group web-producing project in terms of in-depth information, new ways of knowledge acquisition, applications of computer skills, and acquiring many new skills and knowledge beyond the regular classes as well as communication skills.

The data collected from the students’ online journals and interviews further showed that students found the most useful part of the network system provided by the School Cyberfair to be the “works exchange area.” By reviewing web projects of the others, students K01, K02, K03, K04, K05, K06, U04, U09, U10, and H01 were able to: 1) understand the history of the displayed places; 2) learn techniques of website-making from other students; 3) acquire related data or information; and 4) collect recent and useful information about website production, research direction, and data collection.

The online information system was used by the students as a cognitive tool (Jonassen, 2006). The process of using the information system provided by the Cyberfair organization supports Jonassen’s rationales for using technology as cognitive tools. When students work with computer technology, instead of being controlled by it, they enhance the capabilities of the computer, and the computer enhances their thinking and learning.

Throughout the whole process of constructing the project by means of the PBL approach, the winning groups relied heavily on the information system. They kept a daily journal of their progress and wrote their reflexive notes, updated their website, sought problem solving through Q & A, exchanged ideas with other group members, and more. The websites contained abundant resources, databases, related links, and precious experiences and research methods of others that increased students’ interest and motivation and impelled their desire for new knowledge. By assimilating new ideas, they built upon their own intelligence and experiences. In the contest, students also acquired critical-thinking skills through reviewing and critiquing others’ work (U01, K02, K03, & K06). They were able to figure out problems by themselves, and collect and arrange data. By trial and error they developed their own abilities in solving problems.

**Group project links to the curriculum areas**

The significance of the Cyberfair is to encourage school children to connect the knowledge they learn in school to real-world applications. Recognition is given to the outstanding projects in each of eight categories: local leaders, businesses, community organizations, historical landmarks, environment, music, art, and local specialties. Table 5 provides the information about the school code, group project title, category, and links to the school curriculum areas of the four winning teams.

<table>
<thead>
<tr>
<th>School code</th>
<th>Group project title</th>
<th>Category</th>
<th>Link to the curriculum areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 School B</td>
<td>An Aboriginal Tribe in the Clouds of Mt. Bulum LaLa</td>
<td>Environment</td>
<td>Computer/information literacy, mathematics, Chinese</td>
</tr>
<tr>
<td>2 School H</td>
<td>History of Old Fu-wei Street</td>
<td>Environment</td>
<td>Computer/information literacy, social sciences</td>
</tr>
<tr>
<td>3 School U</td>
<td>A half a century of history of the Hon Bridge</td>
<td>Historical landmarks</td>
<td>Computer/information literacy, Chinese</td>
</tr>
<tr>
<td>4 School K</td>
<td>Romance and Scenery of the Islets</td>
<td>Historical landmarks</td>
<td>Computer/information literacy, arts and humanities</td>
</tr>
</tbody>
</table>
The project titles of the four winning groups belong to the two major categories of environment and historical landmarks. Since the Schools Cyberfair encourages students to use technology to share what they have learned and to connect the knowledge they learn in school to the real world, obviously all of the PBL projects commonly link to the subject area of computer/information literacy. Depending on the nature and content of the projects, different projects also have a special link to different curriculum areas, such as mathematics (school B), Chinese (school B & school U), social sciences (school H), and arts and humanities (school K).

School B, for instance, merged mathematics into the web project entitled An Aboriginal Tribe in the Clouds of Mt. Bulum LaLa. In carrying out the PBL project, to introduce the specific living environment of the aboriginal tribe, the teachers had to take the students up to the local divine/giant trees areas in Mt. Bulum LaLa. Interestingly, the students had to apply the measurement skills learned in the class to measure the girth of the giant trees, count the tree rings (Figure 7-1), and estimate the height of the giant trees by applying the similar triangle method (Figure 7-2). They had to collaboratively draw a sketch using ICT. In addition, they used a digital camera or a camcorder to record the PBL process at Mt. Bulum LaLa. Significantly, they actually connected the knowledge they learned in school to the real world and then shared the valuable experience with other students on the Internet through the Schools Cyberfair contest.

![Figure 7-1. Counting a giant tree’s growth rings](image1)
![Figure 7-2. Applying the similar triangle method to estimate the height of the giant trees](image2)

**Conclusion and suggestions**

This study set out to understand, describe, and interpret the uses of information communication technology (ICT) in primary education in the Taiwan Schools Cyberfair as a means to expand and enhance student learning in an extracurricular setting. This study reports on collaborative learning models and describes the effects of applying ICT in project-based learning (PBL), based on the four winning school cases in a competition. Eventually the winning teams will be entitled to participate in the International CyberFair by further translating the whole webpage into English with the assistance of English human resources and additional financial aid from the local educational organizations. Through International CyberFair participation, those students will indeed become community ambassadors on the Internet. They could share self-produced local information with the global community that was otherwise impossible without the PBL activities in the extracurricular learning. The students benefited from the integration of ICT in the learning process and have established international connections.

After this longitudinal study, there are a number of research issues that demand further investigation in the near future:

1. Whether or not those teachers adopt these models in formal educational settings to enhance student learning as suggested above, as long as the students continue to benefit from this type of alternative learning mode, what challenges would they encounter in the exam-oriented school culture in formal education settings?
2. If teachers do not wholly accept our methodology and models, will the students who engaged, directly or indirectly, in our study continue to study in an ICT-supported PBL way?
3. What problems might those students face in an exam-orientated school culture?
4. In which ways might assessment processes take into account less prescriptive ways of learning?
References


Experimental Evaluation of an Instructional Supporting Tool in Distance Learning

Athanasis Karoulis, Ioannis Stamelos and Lefteris Angelis
Dept. of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece // Tel. 0030/2310/99.19.10 // Fax. 0030/2310/99.84.19 // athanasis@karoulis.gr // stamelos@csd.auth.gr // lef@csd.auth.gr

ABSTRACT

This paper describes a controlled experiment concerning the use of a learning aid during an open and distance learning (ODL) course. The core issue of investigation is whether this instructional aid can support, guide, and scaffold the distant student in his/her study. For this purpose, a controlled experiment was conducted with the participation of 191 undergraduate students at the department of informatics at a university in Greece. The considered domain was two lessons concerning human–computer interaction (HCI), the first in usability engineering and the second in interface evaluation methodologies. A test session was also conducted to collect data on the assessment of the effectiveness of the proposed tool. Descriptive statistics as well as a variety of statistical methods have been applied to the collected data in order to test the research hypotheses. The results have shown a statistically significant difference in performance for the student group that used the tool. In order to validate these results, a validity evaluation was also considered and presented. Finally, concerns about the application of the tool in a broader context and further research on the area are also presented.

Keywords
Lesson sheets, Open and distance learning, Controlled experiment

Introduction

Theoretical framework of the study

The constructivist model of learning assumes that knowledge is not transferred but is created by the learner, with the instructor as facilitator (Belanger & Jordan, 2000). Most recent approaches adopt the student-centered approach, which is clear and understandable (Lowman, 1981), is responsive to the ways in which students learn and communicate (Kolb, 1984), acknowledges students’ interests and motivations (Forsyth & McMillan, 1991), and focuses on the explicit needs of learners for meaningful and timely feedback (van Houten, 1980).

All paradigms in this area suggest that ad-hoc prepared instructional material, either in open and distance learning (ODL) or in traditional approaches, has to follow specific guidelines in order to be characterized as sound (Georgiadou & Economides, 2000). Aims and targets, keywords, short chapters and paragraphs, simple language, explanation of difficult points, and exercises and activities are only some of the guidelines. In addition to these, new educational approaches and methodologies have evolved to integrate this material more smoothly into the instructional procedure. Furthermore, various learning aids have been proposed, such as study frameworks, time scheduling, or summaries, because learning issues are, after all, complex and highly case sensitive (Squires & Preece, 1996). So, there is a broad consensus that instructional tools aim to support the student in his/her work, to confront eventual learning difficulties, and to facilitate the students in personalizing the offered information.

Existing research and motivation of this study

In the described context, a new instructional tool, simply called the lesson sheet, has been proposed and thoroughly described in Karoulis, Demitriades, and Pombortsis (2004a). The investigation of the potency of this tool underwent three phases. The first one concerned a qualitative survey, which was applied in traditional classes and was published in the aforementioned study. The second evaluation was a controlled experiment concerning the application of the tool in both a traditional and a hybrid class, and is thoroughly described in Karoulis, Stamelos, Angelis, and Pombortsis (2004b). The work presented here concerns the third phase of the study, namely, to modify the proposed tool to make it suitable for ODL, to apply it in the framework of a pure distance-learning course, and to evaluate its potency and effectiveness in this context. So, a new controlled experiment has been organized and performed in order to elucidate this question.
In short, undergraduate students of the department of informatics studied, through distance learning, two lessons on usability engineering within the framework of a course on software engineering. The students were divided into two groups: a control group, consisting of students who didn’t use the modified lesson sheet, and an experimental group, consisting of students who did use the modified lesson sheet. At the end of the predefined period, an in-person examination took place, with identical questions for both groups. The collected data was analyzed statistically. The results showed a statistically significant difference in the performance of the group of students who used the tool. As the main focus of this paper is to describe the pedagogical findings and to raise discussion based on these findings, this result points towards the application of a simple but valuable tool for supporting, guiding, and scaffolding the distant student during his work at home.

We emphasize at this point that prior pedagogical findings of the application of the tool in traditional and hybrid (blended learning) settings have already been published in Karoulis, Demetriades, & Pombortsis (2004) and Karoulis, Stamelos, Angelis, & Pombortsis (2004) so present work aims to add the findings of its application in ODL settings, as subsequently described. In order to be precise in this evaluation, we followed a rigorous methodological framework in this paper. While this approach is encountered often in other disciplines, it is greatly neglected in contemporary ODL research.

Lesson sheet

The proposed tool has its origins in the domain of open learning. In table format, it is divided into rows and columns. The left-hand column contains the outline of the lesson in paragraph form, while the right-hand column provides a variety of context-sensitive information that could be a summary of the subject under consideration, such as charts and graphs, pictures, or other material related to the discussed issues. In addition, adequate free space allows students to personalize the sheet with his/her own notes — information that should have meaning to students when they work at home.

The basic construction guidelines of the lesson sheet are listed below:

- The lesson sheet uses the shape of a table to frame and structure the offered information.
- The left-hand side contains, the information presented. This provides an outline for every lesson.
- There must be adequate recall of situated knowledge: an already known definition, image, graph, etc.
- On the right-hand side there is space for the student to note whatever he/she wants relevant to the corresponding lesson to thus personalize the lesson.
- Images, figures, charts, and any other appropriate means are included to help students visualize the information.
- Notes that have nothing to do with personalization, such as Internet links, tables, and additional or external information and resources, are provided on the sheet in order to accelerate the procedure and diminish the distraction that occurs during the students’ note-taking time.
- The left-hand side includes the necessary self-assessment exercises and the activities that the student has to perform during the lesson, while the corresponding space on the right is usually reserved for the answer. Exercises gradually increase in difficulty.
- It must provide feedback on the self-assessment exercises so that the student can correct his/her mistakes, by various means, such as pinpointing the relevant sections on the instructional material or suggesting further readings or external sources.
- Lesson sheets contain text and theory of no more than 2 to 3 paragraphs in length for every section. In addition, an alternative option such as an exercise, an external resource, or a student activity is provided.
- The use of the sheet in practice has shown that 4 to 5 pages are sufficient for a 2-hour lesson. In addition, good image and print quality are important.

Appendix A is a simplified sample lesson sheet, which clarifies the basic structure. The sample is based on the lesson sheet used during this experiment.

Experiment

Since this study concerns a controlled experiment, emphasis was given to two areas. First, experimental settings were made as realistic as possible with respect to those in practice. To fulfill this requirement, the lesson sheets were
applied to real university-level, undergraduate lessons, delivered through distance learning. Furthermore, an external motivation existed because the test score counted towards each student’s final score. In this way, the students perceived this series of sessions to be part of their syllabus. Second, planning, operation, and analysis were done according to the methods proposed in the literature, such as in Wohlin et al. (2000). Consequently, the procedure described adheres to these guidelines.

**Definition**

To define the experiment, the goal-question-metric model by Basili & Weiss (1984) and Basili & Rombach (1988) was used. The formal definition of this experiment is as follows:

**Analyze**

For the purpose of

With respect to their

From the point of view of

In the context of

**the application of the lesson sheets**

**assessment**

**effectiveness during the instructional procedure**

**instructors and instructional designers**

**distance learning university level lessons on usability engineering**

**Planning**

The context of the experiment is determined in this section. This includes personnel and the environment, the selection of variables, measurement scales, the statement of the hypotheses, and so on.

**Context selection**

To achieve the most general results, the experiment should be executed in real settings. However, due to various constraints, this is often not feasible, so a cheaper alternative is chosen, such as using a restricted version of the real environment or students as experiment subjects. Such approaches are cheaper and easier to control, yet they seldom address real problems and are more directed to a certain context. So, the context of the experiment must be defined according to following four dimensions (Wohlin et al., 2000):

- **Online vs. off-line.** Online, since it was performed during the semester’s lessons.
- **Student vs. professional.** Students are the final user population to use the tool, so “students” and “professionals” in this context are equal.
- **Pretend vs. real problem.** The experiment addresses a real problem under real circumstances (real lessons).
- **Specific vs. general.** This investigation concerns a specific problem, namely, the subject of usability engineering. Generalization to other domains may be made (if at all) after having considered the external validity of the experiment. This issue will be addressed later.

**Selection of variables**

The independent variable (factor) is the use of the lesson sheet. It is of nominal type, with two possible values (yes & no).

The dependent variable is the students’ performance on the taught domain, measured in terms of a grade on a test. This variable is measured in ratio scale, from 0 to 150, and is operationalized in terms of 15 questions depicted in the examination sheets of the final test.

**Null hypothesis**

\( H_0: \text{There is no difference in the students’ performance on the taught domain due to the use of the lesson sheets.} \)

The alternative hypothesis is consequently:

\( H_1: \text{There is a difference in the students’ performance on the taught domain due to the use of the lesson sheets.} \)
Selection of subjects

Two independent groups were formed from 191 participating students attending the fifth semester at the department of Informatics at the Aristotle University of Thessaloniki, Greece (AUTH). The students were all volunteers, and the grade they would acquire at the scheduled final test would count towards a bonus of up to 10 percent above their final grade. They were divided into a control group of 96 students and a treatment group of 95 students. Students had been previously grouped based upon their choice of percentage in the following statement, which appeared in the header of examination document: “Percentage of the use of the tool during your study: 0%, 10% … 90%, 100%.” According to their declaration, students who had used the sheet less than or equal to 50 percent were classified as the control group, whereas students who had used it more than 50 percent were classified as the treatment group. The final classification was balanced (although this was not intended): 96 vs. 95 students.

Educational framework

Initially, the supporting website was set up. Through this site, the following educational material, consisting of four main learning files and some supporting material, was delivered to the students:

- The lesson sheet, a file named “Guide.pdf”
- The two lessons files, “HCI01.pdf” and “HCI02.pdf”
- A third learning file, “Presentation.zip,” an Authorware file. As stated to the students, “this would be the file a traditional teacher would use to present the material, if these lessons were in traditional form.”
- A shareware book (Task-Centered User Interface Design, by Lewis & Rieman, 1994) and some relevant Internet resources (provided as additional, voluntary instructional material).

Experimental design

Lessons

The students participated in, over distance, the equivalent of two 3-hour lectures (one per week) concerning two selected aspects of usability engineering (UE): introduction/definitions of UE, and evaluation.

The studying material consisted of two pdf files, one for each “lecture.” There was also a multimedia presentation, presenting all aspects the text files in a multimedia way. External resources, such as additional text books and relevant websites were also provided. The students had also the option to contact a mentor (one of the authors) if they had questions regarding their study. In conclusion, the design aimed to imitate a self study in a distance environment, corresponding to two real lessons and providing adequate support to the student.

The main supporting and guiding file was the modified and enriched lesson sheet, also provided at the supporting website. The students had the choice of whether to use it or not. They only had to declare this percentage of usage at the final test, in order to be classified into a group, as already described.

Tool

The lesson sheet was a detailed guide that accompanied students through their study, pinpointing pages to read, providing self-assessment exercises (with answers at the end of the sheet), encouraging them to visit the supporting Internet resources (where relevant), providing explanations on unclear or difficult aspects of the material, and providing pictures or diagrams where appropriate. In other words, the sheet was designed to be scaffolding and a guiding factor for the student who would rely on it. Obviously, this last claim is the main one to be proved through this experiment.

Examination

The final test took 2 hours and consisted of an examination sheet with 15 multiple choice questions to facilitate the statistical analysis.

The experimental groups were, as already stated, randomized and almost balanced with 96 and 95 participants respectively. No blocking of any subject with respect to other factors was made.
The purpose of this design was to depict and measure the impact of the learning sheet on the students’ performance. In this case, it was most appropriately recorded through a written test, as there were a large number of participants, which prohibited other approaches such as interviews or oral examinations. Note that there were no criteria such as success or failure for the students, only the acquired grade in a scale from 0 to 150, as there were 15 questions in the examination sheet, each worth 10 points.

Other materials and information
In addition to the lesson sheets described above, students were also supplied with the following:
- Two distance lectures, as Adobe pdf files, which concerned UE basics and evaluation methodologies. These lectures were chosen because they provide many new notions and include simple as well as complex material. Students in both groups received the same quality and quantity of information.
- Instructions to students on the website, the time scheduling etc. It was emphasized that the accompanying lesson sheet was a valuable, yet optional, aid in their study. However, as results showed, only half of the students used it in the way it was intended to be used.
- A presentation of the issues discussed in the pdf files. The Authorware multimedia-authoring tool was preferred here over PowerPoint because of its augmented multimedia capabilities.
- Additional studying material, such as a shareware book and some Internet resources, as already described.
- The final examination.

Guidelines to subjects
The students were given no special guidelines, apart from those provided on the website. The initial study of the tool (Karoulis, Demetriades, & Pombortsis, 2004) showed that it was usable and its use quickly became transparent, so no extra preparations had to be made at that point. The results of this study, concerning the ease of use of the sheet, confirmed those of the aforementioned study.

Measurement Instruments
The main means of depicting the students’ performance was the final examination, which consisted of 15 questions, covering all aspects discussed in the pdf files. An extra percentage scale, from 0 percent to 100 percent, was added at the header of the document, where the students pinpointed the usage of the lesson sheet during their study.

The questions were of quantitative type in order to be easily analyzed statistically. Five kinds were employed in this experiment:
- Multiple choice: one of several correct answers is possible
- Yes/no or right/wrong: bipolar differentiated notions
- Choose from a list: more than one can be correct
- Matching: pairing corresponding items from two columns
- Fill in the blank: the student must fill in the correct word

Appendix B shows a sample of the examination sheet used (one page out of four) presenting some of the questions used.

Experiment
Preparation
The preparation of the experiment underwent the following steps, each corresponding to the experimental components:
- Subjects. The students were asked to express their interest in participating in the experiment by having personal contact with one of the tutors or communication via e-mail.
- Material preparation. The lesson framework (what to teach), was defined; the methodology (how to teach) was chosen; the lesson sheet and other files (by what means) were prepared prior to the series of lessons; and the support site (for use during the lessons) was set up.
- Instruments. What had to be examined was decided. Once the pool of questions had been defined, the questions to appear on the examination sheet were chosen.
The procedure described here took one and a half months. Only after the final test were the subjects informed in detail via the website about the experiment in which they took part and its basic structure, yet they knew from the beginning that research would also, in parallel, take place.

**Execution**

The execution of the experiment was as predicted. There was no deviation from the plan. All students downloaded the study files and kept the schedule. It must be emphasized at this point, that the final test was performed under real circumstances, in separate rooms with the aid of three additional overseers.

Finally, a statistical analysis, as subsequently described, was applied in order to investigate any significant difference in the performance of the two student groups.

**Data validation**

The data collected through the examination sheets was considered reasonable, as there was no indication to the contrary. Data was also considered to have been collected correctly, as it was recorded on the examination sheets, and the subjects were considered serious (applied correct treatments in correct order), as they were well motivated.

**Data analysis & interpretation**

After the separation of the students into both groups, the statistical analysis of their performance was made. Descriptive statistics showed that the mean values of the two groups already provide an obvious differentiation, 2.84 to 5.39, nearly double. Standard deviation had a relative low value in both cases (1.23 and 1.34), which means that the students’ grades were more or less concentrated around the mean values. Table 1 provides these facts.

*Table 1. Descriptive statistics of control and treatment group*

<table>
<thead>
<tr>
<th>Group Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of the tool</td>
</tr>
<tr>
<td>No (&lt; 50%)</td>
</tr>
<tr>
<td>Yes (≥ 50%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of the tool</td>
</tr>
<tr>
<td>Grade No (&lt; 50%)</td>
</tr>
<tr>
<td>Yes (≥ 50%)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

The data of table 1 is depicted graphically in the box-plot and bar graphs in Figure 1.

*Figure 1. Box-plot (a) and bar graph (b) of the mean values of the grades of the students*
This statistically significant differentiation, visible in these graphs, had to be investigated with a series of inferential statistical tests, which are subsequently presented.

Initially, a $t$-test showed a statistically significant differentiation in a statistical significance level of $p < 0.0005$, which is unusually low, as shown in Table 2 below.

<table>
<thead>
<tr>
<th>Table 2. Results of the t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test of independent samples</strong></td>
</tr>
</tbody>
</table>
| \begin{tabular}{|l|c|c|c|c|c|c|c|c|}
| \hline
| Levene’s test for equality of variances & F & Sig. & $t$ & df & Sig. (2-tailed) & Mean Difference & Std. Error Difference & 95\% Confidence Interval of the Difference \hline
| Equal variances assumed & .256 & .608 & -9.460 & 94.000 & .005 & -2.54611 & .26915 & -3.08051 \ -2.01171 \hline
| Equal variances not assumed & & & -9.251 & 68.724 & .005 & -2.54611 & .27523 & -3.09521 \ -1.99701 \hline
| \end{tabular} |

A comparable result also came from the non-parametric Mann-Whitney test (significance level $p < 0.0005$) as well as the Wilcoxon-W test, which were both applied to strengthen this result.

<table>
<thead>
<tr>
<th>Table 3. Non-parametric Mann-Whitney and Wilcoxon W tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test statistics (^a)</strong></td>
</tr>
</tbody>
</table>
| \begin{tabular}{|l|c|}
| Mann-Whitney & 186.000 \hline
| Wilcoxon W & 2016.000 \hline
| Z & -6.766 \hline
| Asym. Sig. (2-tailed) & .005 \hline
| \end{tabular} |

\(^a\) Grouping variable: tool usage

![Figure 2. Scatterplot of correlation between grade and usage of the tool](image-url)
A new question emerged here, namely, whether the usage of the lesson sheet had an impact on the final grade of the student. In other words, did frequent use of the lesson sheet improve the student’s score on the final test? This statement obviously enhances the stated research hypotheses; however, it was not initially set.

Because of this, the researchers had to go a step further. First, a dispersion scatterplot was calculated, shown in Figure 2. A linear correlation is apparent. To prove this claim, the Pearson-\( r \) correlation coefficient has been employed.

Table 4, shows that the Pearson-\( r \) (0.756) is statistically significant at the 0.01 level and the determinant coefficient (\( r^2 = 0.572 \)) shows that a significant part of the dispersion of the grades (57.2%, shown in Figure 2) can be statistically explained as caused by the use of the lesson sheet.

<table>
<thead>
<tr>
<th>% tool usage</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>% Tool usage</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>% tool usage</td>
<td></td>
<td>.756**</td>
<td>.000</td>
<td>1</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td>.000</td>
<td>.96</td>
<td>96</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed)

**Interpretation**

It is obvious that the group that claimed to have used the tool more than 50 percent performed significantly better than the other group. This result is clearly depicted in the box-plots.

**Hypotheses answering**

So, the stated hypotheses can now be answered. According to what has already been discussed, the null hypothesis \( H_0 \) must be rejected and the alternative \( H_a \) adopted instead. Use of the lesson sheets does affect students’ performance on the test.

However, this difference could always be the result of a third factor causing this effect (causal relation). In other words, the “good” students could very well be the ones that really used the tool and benefited from its impact. This is a serious consideration and will be discussed later.

**Generalization**

Generalization can only be made in the context set by the particular experiment. However, Wohlin et al. (2000, p. 112–113) provide a good discussion on conclusions that can be drawn and generalizations that can be made in a broader context.

To make an attempt for any generalization, one must consider the external validity of the experimental settings. As discussed in the next section, the experiment was performed in real circumstances, so subjects, settings, and history do not threaten generalization. Therefore, it can be argued that the results of this study can be generalized in the context of university-level lessons. However, there is a concern regarding the cognitive domain to be taught, so this generalization must be considered as tentative, as the particular cognitive domain is an important factor.

**Practical importance**

The practical importance of the results presented here could be of interest to anyone dealing with ODL. Educational designers especially are seeking any aid in the direction to support, scaffold, and guide the distance-education student, so this proposed tool, even in its printed form, could provide a fair solution, as shown by the studies so far.
As for a potential replication of the experiment, the described context can be easily mimicked, so a replication of this experiment is always possible.

Validity evaluation

The validity of the conclusions is highly dependent on how well specific threats were manipulated. According to Cook and Campbell (1979), validity is subdivided into four types, each of which addresses a specific methodological question and includes one type of threat to the validity of the results. The four threats are conclusion, internal, construct, and external validity. There is a strong relationship between these four validity threats (Trochim, 1999; Wohlin, 2000, p. 64), depicted in Figure 3.

![Figure 3. Relationships in the experiment design with respect to validity. Adapted from Trochim (1999)](image)

The main idea behind this relationship is that, in order to draw valid conclusions about the theory defined in the hypotheses, based on observations, one must provide satisfactory answers to all the specific methodological questions that evolve. So, in this section, the four validity threats are discussed as well as the approaches employed to confront them. However, aspects that are irrelevant to the particular context, or obviously not applicable, were neglected.

Conclusion validity

Conclusion validity, point 1 in Figure 3, is concerned with ensuring that there is a statistical relationship between treatment and outcome. So, threats to conclusion validity are factors that can lead to incorrect conclusions about relations between the treatment and the outcome of the experiment. According to Cook and Campbell (1979), factors such as low statistical power or violated assumption of statistical tests threaten conclusion validity. In this case, the reliability of the measures is ensured by the use of the examination sheets. So, as this approach constitutes an objective rather than subjective measurement mode, only unintentional errors could occur during the grading phase. Because the experiment occurred under real circumstances, during a university-level lesson, the reliability of treatment implementation was also ensured, as all subjects were treated in a similar manner.

Internal validity

Internal validity, point 2 in Figure 3, means that the observed relationship is a result of the factor (i.e., causal). Threats to internal validity are influences that can affect the independent variable with respect to causality, that is, they can affect the conclusion of a possible causal relationship between treatment and outcome. According to Cook and Campbell (1979), factors such as fatigue effects (negative), learning effects (positive), or instrumentation may threaten internal validity. Although single group threats of internal validity are considered to affect only experiments with no control group to which the treatment is applied (Wohlin et al., 2000, p. 69), some of these factors may affect this experiment as well. However, the quantitative approach followed to assess the performance of
the students diminishes the threat of an eventually suboptimal instrumentation. Regarding the selection of subjects, volunteers are considered to be generally more motivated (Wohlin et al., 2000, p. 69), as it was in this case.

**Construct validity**

Construct validity, point 3 in Figure 3, concerns the generalization of the experimental result to concept or theory behind the experiment. So, if the relationship between cause and effect is causal, then treatment reflects cause and outcome reflects effect.

_Inadequate pre-operational explication of constructs:_ The main concern in the described experiment is whether an examination sheet really depicts the acquisition of offered information, and what exactly is measured by the examination sheet. To confront these aspects, the examination sheets have been designed according to two factors. Firstly they completely covered the taught domain: seven questions concerned the first lesson (usability engineering — design) and eight questions concerned the second lesson (usability engineering — evaluation). So it can be argued that the acquisition of the offered information must be complete and broad on the domain, in order for the subject to perform well on the test. Secondly, the quantitative questions aim to measure this acquisition in a clear scale from 1 to 10 for every question, giving a total of 150 points. So, it can be argued that the acquisition of the information can also be measured, yet the assimilation of the information cannot. The passing from memorization to learning and understanding is a complicated issue, and was outside the scope of this study. However, it is discussed in Karoulis, Demetriades, & Pombortsis (2004). To conclude, the described design of the examination sheet aimed to confront this specific threat as well.

_Mono-operation and mono-method bias:_ As the experiment concerned only one domain (usability engineering) and used only one measurement instrument (the examination sheet), this threat was present. The strict design and conduct of the experiment diminishes these threats.

**External validity**

If there is a causal relationship between cause and effect, can it be generalized? Therefore, external validity, point 4 in Figure 3, is related to generalizing. By making the experimental environment as realistic as possible, threats to external validity are reduced and research results can be both generalized to the population under study as well as to other research settings.

According to Cook and Campbell (1979), interactions between selection, settings, history, and treatment may threaten external validity. However, in this case the real settings of the experiment diminish most of those threats as well.

**Discussion and concerns**

However, the greatest threat to the presented validity evaluation was omitted, on purpose. It will now be discussed in detail. One point in Conclusion Validity (random heterogeneity of subjects) and one point in Internal Validity (selection and randomization) cannot be considered as addressed by the experiment design and remain as validity threats. In more detail, a very plausible assumption could well be that only the “good” students adhered to the guidelines provided by the lesson sheet and really used the tool as it was proposed; therefore, it is a reasonable conclusion to assume that they performed better in the test. The results from the first study on the lesson sheets, where only the good students accepted the tool, while the rest ignored it, strengthen this claim.

However, there are some reasons to believe that this is not as bad as it seems. Firstly, education in general underlies this threat, so why not ODL? In every school a number of pupils and students try to cheat, do not follow instructions and are accordingly characterized as “bad.” This is a social phenomenon that is completely outside of the scope of this study.

Secondly, ODL presumes the personal responsibility of the, usually adult, student. The inactivity, loss of interest, and finally dropout phenomenon is a well-known problem in ODL. This is a main concern in this study. The distance
student needs support, guidance, and scaffolding, elements that usually are not present in the instructional material. Many studies up to now, such as Bernath and Rubin (2004), Pierrakeas, Xenos, Panagiotakopoulos, and Vergidis, (2004), and Xenos, Pierrakeas, and Pintelas (2002), pinpoint that the drop-out rates vary between 25 percent and 30 percent, in a roughly estimated mean value, as it strongly depends on the course and educational context. From this point of view, the encouraging results of the present study argue for the use of the tool in more cases, because only a broad application and a number of reports on its effectiveness can clarify its potency.

Another concern of the present study, not visible in the validity evaluation, is the construction of the educational environment. Contemporary distance education separates the instructor from the students in time and space, yet provides remarkable educational results outside the context of the traditional class (Keegan, 1996). It is also argued that the use of enriched multimedia material provides a better instructional environment and augments the quality of the delivered education (Owoc, Maciaszek, & Hauke, 2001). In our case, these attributes were present, specifically the printed educational material combined with the multimedia presentation. However, it is debatable if such a poor educational setting constitutes a real ODL environment. There were, however, some specific reasons for this approach. Firstly, the validity of the experiment had to be ensured, so a very simplified ODL environment was favored over a more innovative one. Although this reason could be enough, as it confronts many validity questions, it was also argued that the potency of a new proposed tool should also be examined in an environment familiar to the subjects, to avoid any special adaptation training that could interact with the experiment settings. Such a “minimalist” approach has also been reported to perform adequately (Karoulis, Sfetsos, Stamelos, Angelis, & Pombortsis, 2004), so the present study could focus untrammeled on the effectiveness of the lesson sheets.

However, despite these claims, the question of the validity of the results of this study in a real ODL environment (with tutors, tutorial sessions, delivered over the Internet, etc.) remains unanswered. It is obvious that more research has to be undertaken on this aspect. However, an aid comes through Wedemeyer (1973, p. 101), who defines the notion of “freedom of the distant student” according to three parameters:
1. The learner must be able to define his own pace of learning
2. Learning must be individual and personalisable
3. The learner must be able to choose his instructional targets as well as the means to achieve them.

The point of discussion here is in how far the lesson sheet can assist the distant student in these directions of ODL, both in the presented study, as well as in a more traditional ODL setting. It can be argued that this tool, which can be ad-hoc constructed for every educational context, adheres to these pillars. The basic construction guidelines, presented in section 2, more or less support all three ODL principles stated by Wedemeyer (1973). Former studies, as well as this study, strengthen this claim. However, there are always doubts left, due to the variability of the ODL settings, and the diversity of the entities employed and interacting during any instructional process.

Another concern affects the practical importance of the proposed tool, as already discussed, and its application independent of the cognitive domain. Up to now, the tool has been applied in six different settings. Four were at the post-graduate level and six at the under-graduate level. Although a generalization can hardly be attempted, indications show a common acceptance of the tool by all classes, especially according to qualitative approaches such as interviews. Therefore, more study on multiple domains is needed to draw a valid conclusion regarding broad application of the proposed tool.

**Further research**

Further research should also consider the transition of the tool to a more technology-based format, such as an interface agent or a browser plug-in. URL resources, animated diagrams, or adaptive interactions with the student are simple possibilities of this approach. The necessary alterations should be investigated, and some formal testing of the adapted tool will provide valuable information on its use in sensitive aspects, such as the scaffolding and support for students who are studying at a distance. From a technological perspective, if the tool is successfully implemented in ODL, it could greatly aid in solving the problem student dropout, which is the number one threat in almost all ODL programs.
Conclusion

The underlying pedagogy of the lesson sheets has already been published in three prior works. This paper deals only with its evaluation in a distance-learning setting and proposes a more rigorous methodology to do this, based on an extensive statistical elaboration. Although statistical findings do not add anything from a pedagogical perspective in the study, they prove better performance of the group that used the tool.

Based on the aforementioned results, it can be concluded that the lesson sheet, in its present form, can provide a valuable instructional aid in cases of simple ODL environments and an indication of its effectiveness in more complex environments as well.

This conclusion should be kept in mind when it comes to the design of the sheets for application during the instructional procedure. However, it is of paramount importance to always keep in mind the basic construction guidelines, presented in section 2, and adapt them according to the particular educational context.

References


P ierrakeas, Ch., Xenos, M., Panagiotakopoulos, Ch., & Vergidis, D. (2004). A comparative study of dropout rates and causes for two different distance education courses. *International Review of Research in Open and Distance Learning, 5*(2).


Appendix A: A sample lesson sheet

**Learning Guide**

Use this learning guide as a support to your study. Adhere to the flow of this guide and follow the instructions. You will use the learning material that you have downloaded. It is important to elaborate the proposed exercises and to work with the additional material, as pinpointed by this guide. You can make notes on this sheet, thus personalizing it to your educational demands. It will not be returned to the instructor but will become a study source and reference for your own use.

**Part A: Usability Engineering**

<table>
<thead>
<tr>
<th>1. Introduction — the field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface (pages 19–21 on the HCI01.pdf file that you have downloaded) (Explain the differences — Note the important aspects)</td>
</tr>
<tr>
<td>Interface: Interaction: Interfere:</td>
</tr>
<tr>
<td>The picture by Norman K. (page 21) (Study it for a couple of minutes and make comments on it on the sheet)</td>
</tr>
<tr>
<td>Exercise 1: In a few words, explain the presented terms in the interaction schema above, as well as their contribution to the interaction Write your answer here →</td>
</tr>
<tr>
<td>Interface as “Inter-face” (page 22)</td>
</tr>
<tr>
<td><img src="image" alt="Interface as “Inter-face”" /></td>
</tr>
<tr>
<td><strong>ATTENTION! NOT in your handbook!</strong></td>
</tr>
<tr>
<td>Definition Interface is...</td>
</tr>
<tr>
<td>Other definitions:</td>
</tr>
<tr>
<td>• Preece: the sides of the system...</td>
</tr>
<tr>
<td>• Shneiderman:</td>
</tr>
<tr>
<td>Remember to enrich this sheet with your own comments, notes, suggestions, etc. Your final review will be a passage through this sheet.</td>
</tr>
<tr>
<td>A figure for this definition (page 23 → read your book AFTER you have studied this figure and AFTER you have performed the exercise below.)</td>
</tr>
<tr>
<td><img src="image" alt="A figure for this definition" /></td>
</tr>
<tr>
<td>Exercise 2: a) Describe aloud the human-system interaction according to this figure (start at “demand”) b) What does the grey area mean? (The answer is at the end of this sheet.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Man–Machine Interaction and Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions (page 24) →</td>
</tr>
<tr>
<td>Exercise 3: Mark the corresponding terms Fill in the blanks →</td>
</tr>
<tr>
<td>Learning ease</td>
</tr>
<tr>
<td>Low error rate</td>
</tr>
<tr>
<td>Remembering</td>
</tr>
<tr>
<td>The answer is provided at the end of the sheet</td>
</tr>
<tr>
<td>Additional stuff</td>
</tr>
</tbody>
</table>
Appendix B: A Questionnaire sample

Test in Usability Engineering

Name: ____________________________ ImNr: _______ Date: ______/____/200__
Score: __________________________

Questions
(Please note an X only where appropriate)

1. According to ACM, the fundamental domains of Informatics are:
   - [__] Hardware
   - [__] Windows
   - [__] Artificial intelligence and robotics
   - [__] Distributed systems
   - [__] Object-oriented programming
   - [__] Databases and information retrieval
   - [__] Human–computer interaction
   - [__] Data structures
   - [__] Communications systems
   - [__] Logical programming
   - [__] C language and UNIX
   - [__] Digital electronics
   - [__] Software engineering

2. Which one is the correct Human-computer communicative interaction?
   - [__] Demand – Coding – Interpretation – Processing – Presentation – Perception – Evaluation

3. At the definition of ISO: “Usability = Effectiveness + Efficiency + Satisfaction” define:

   Effectiveness = [__] Efficiency + Accuracy
   - [__] Accuracy + Completeness
   - [__] Utility + Accuracy
   - [__] Efficiency / Resources
   - [__] Validity + Accuracy
   - [__] Accuracy / Validity
   - [__] Validity / Accuracy
   - [__] Validity / Resources

   Satisfaction = [__] Positive Feeling + Positive Attitudes
   - [__] Positive Feeling / Positive Attitudes
   - [__] Happiness + Positive Attitudes
   - [__] Comfortability + Positive Attitudes
   - [__] Happiness + Positive Feeling
   - [__] Comfortability + Validity
   - [__] Utility + Validity
   - [__] Positive Feeling + Positive Attitudes

4. Which are the usability parameters, according to Nielsen?
   - [__] Efficiency of interaction
   - [__] Satisfactory feeling
   - [__] Subjective user satisfaction
   - [__] Error protection
   - [__] High work efficiency
   - [__] High user efficiency
   - [__] Efficient work
   - [__] Satisfactory system efficiency
   - [__] Low user error rate
   - [__] Clear file structure
   - [__] Navigation speed
   - [__] Easy to learn
   - [__] Easy to remember the use of the system
   - [__] Easily recovery from errors
   - [__] Easy navigation

5. Which of the following is correct for the design models?
   - [__] The waterfall model demands resource definition.
   - [__] The waterfall model includes requirement analysis, design, implementation, and evaluation.
   - [__] The waterfall model is iterative.
   - [__] The waterfall model is based on evaluation.
   - [__] The waterfall model demands a working team.
   - [__] The star model functions after the completion of the system
   - [__] The star model doesn’t include the implementation phase
   - [__] The spiralette is iterative
   - [__] The spiralette is user-centered
   - [__] The spiralette is based on evaluation
   - [__] The spiralette demands resource definition
   - [__] The spiralette is applied after the completion of the system
Beyond Sharing: Engaging Students in Cooperative and Competitive Active Learning

Gloria Yi-Ming Kao
Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C. //
gis92801@cis.nctu.edu.tw

Sunny S. J. Lin
Institute of Education, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C. //
sunnylin@faculty.nctu.edu.tw

Chuen-Tsai Sun
Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C. //
ctsun@cis.nctu.edu.tw

ABSTRACT
The authors describe their design for an Internet-based learning environment called BeyondShare in which students are encouraged to gain a deep understanding of the learning material, reflect on the quality of individual constructions through sharing and peer evaluation, and synthesize cross-unit knowledge by integrating self- and peer-produced constructions. A sharing typology may consist of basic sharing, sharing with notification, sharing with feedback, or sharing with interaction. BeyondShare distinguishes itself by combining sharing activities with task structuring and cooperation/competition to achieve active learning. A formal evaluation of BeyondShare was performed with a class of 34 college students who created concept maps for a computer programming language and who were encouraged to become active learners by exchanging roles throughout the experiment. Evaluation results indicate that the students viewed BeyondShare as an easy-to-use environment that motivated them toward comprehensive knowledge integration by sharing construction products with their peers. Potential activities and suggested modifications are discussed.

Keywords
Active learning, Cooperation, Competition, Distributed sharing construction, CSCL environment

Introduction
The concept of sharing has taken on new importance in a world that has the Internet — a tool that allows for resource access from any place at any time. Examples of Internet-based sharing include personal websites, blogs, discussion forums, and instant messaging; a growing number of applications support sharing using different media (e.g., del.icio.us, Flickr, and YouTube). These tools disseminate individual or group beliefs in a manner that binds geographically dispersed individuals with common interests. When applied to group-based pedagogy, the anyplace-anytime characteristic enables a shift from real-time learning to asynchronous distributed learning (Kreijns, Kirschner, & Jochems, 2002). The same characteristic enables researchers to create sharing activities that entail concurrent, multi-user interactions (Greenberg & Marwood, 1994; Yang, Chen, & Shao, 2004). One example is the use of information technology tools to share musical ideas via exchanges of audio files instead of through verbal discussions of concepts (McCarthy, Bligh, Jennings, & Tangney, 2005).

However, many pedagogical or research projects address the how or what of sharing to benefit collaborative learning without questioning the why or examining the effects of sharing on learning contexts. To reap the benefits of collaboration entailing mutual engagement as opposed to simple cooperation entailing labor divisions (Roschelle & Teasley, 1995), teachers and researchers frequently design tasks that involve information sharing followed by discussion (see, for example, Häkkinen, Järvelä, & Mäkitalo, 2003). The interactive structure of computer-supported collaborative learning (CSCL) environments creates additional constraints or freedom for learners. One of several impediments to a desired social interaction is the tendency to assume that it will automatically occur because the environment makes it possible (Baker, Hansen, Joiner, & Traum, 1999; Kreijns et al., 2002). Research suggests that few students are willing to participate in CSCL discussion forums without some additional motivation, and that
factors such as social loafing (e.g., the “free-rider” and “sucker” effects) can lead to responsibility diffusion (Barron, Kerr, & Miller, 1992). Consequently, spaces set aside for collaboration or cooperation are often misused for chatting or storage at the expense of the desired goal of collaborative learning through sharing.

Such discrepancies may be due to a lack of sufficient structure — for instance, the failure of teachers to completely organize learning tasks. We addressed this issue by viewing sharing as an intermediate step in a process consisting of active engagement in meaningful learning and knowledge integration. Specifically, learning roles are made more active and meaningful as students a) construct personal concept maps for an assigned learning unit, b) share personal concept maps across units while critically evaluating their peers’ contributions from other units, and c) actively integrate concept maps across all units using a meta-plan to create a “patchwork” of knowledge. Process details will be described in a later section.

In other words, our BeyondShare approach emphasizes the integration of cross-unit knowledge in pursuit of personal goals to generate productive exchanges among students. Instead of expecting students to automatically share resources and negotiate with each other in a CSCL environment, we injected a sense of competition to encourage active learning. As part of this sharing process, we introduced a cooperative competitive learning (CCL) strategy (Lin, Sun, & Kao, 2002) that accommodates both cooperation and competition in a manner that yields greater intrinsic motivation (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Tauer & Harackiewicz, 2004). Our formal evaluation of BeyondShare was designed to answer the following research questions:
1. How many students are able to finish “beyond-sharing activities” (to be described in a later section) using BeyondShare?
2. Did students perceive BeyondShare as easy to use?
3. Did the three activities designed for BeyondShare evaluation achieve the goal of promoting active learning?
4. What percentage of students became actively engaged in both personal and sharing construction?
5. Did a larger percentage of students engage in active learning during personal construction or sharing construction activities?

**Background**

**Sharing**

A considerable amount of research in this area has focused on building a shared sense of understanding or meaning, that is, finding common ground within groups in collaborative learning settings (e.g., Baker, Hansen, Joiner, & Traum, 1999; Mulder, Swaaj, & Kessels, 2004). Four categories can be created according to this perspective (items 1–4, Fig. 1; black silhouettes represent students who play active roles):

1. Basic sharing. Citing or using an idea from a peer is the most basic sharing format. However, most learning situations lack proper motivation for sharing; therefore, some self-regulated individuals model or cite works while others do not, even when requested or instructed. Furthermore, those who benefit from sharing usually have no channel for notifying idea originators, who therefore remain unaware of how others use their ideas.
2. Sharing with notification. In this variation of basic sharing, cited authors are notified that their ideas are being used. Various technologies, such as Really Simple Syndication (RSS), allow authors to push their latest ideas to subscribers, thus facilitating the timely spread of knowledge.
3. Sharing with feedback. By providing feedback, users help the original authors revise and improve their work. The Computer-Supported Intentional Learning Environment (CSILE) constructed by Scardamalia and Bereiter (1991) is one example of a method designed to promote user feedback.
4. Sharing with interactions. Authors can interact via discussion threads, for example, Greenberg and Marwood’s (1994) GROUPKIT (see also Yang et al., 2004). However, participation requires individual motivation.

Researchers such as Häkkinen et al. (2003) and Mulder and Swaak (2002) have used qualitative, quantitative, or a combination of the two approaches to assess collaboration during the sharing process. Completed acts of sharing are followed by quality discussions. Special attention must be paid to the effects of group dynamics on shaping shared meaning (Stahl, 2005), while acknowledging that shared contributions cannot be accepted as indicators of shared understanding among all team members (Beers, Boshuizen, Kirschner, & Gijselaers, 2005). In other words, it is important to separate the term shared knowledge (Edmonds & Pusch, 2002) from shared understanding or shared
meaning. While researchers expect to bring shared understanding into full play in a collaborative learning context, they must note whether the learning activities are structured in a manner that facilitates mutual understanding rather than simple exchanges of information.

Today’s Web 2.0 (O’Reilly, 2005) technologies facilitate different applications (e.g., blogs, Wikipedia, del.icio.us, Flickr, YouTube) that support the sharing of various kinds of multimedia content. These applications are popular because users enjoy expressing their own viewpoints by distributing their articles, bookmark collections, photos, or video clips, and readers/viewers enjoy or use the information gathered from the shared works. These applications all have the same key element: providing users with spaces to share their work and/or to find others with similar interests. In other words, to some degree they all fit into one or more sharing typology categories. For example, most bloggers are interested in sharing hyperlinks with others interested in the same domain knowledge, yet bloggers in the same domain may compete to attract more visitors to their blogs and therefore work to maintain a favorable page ranking on a major search engine. This phenomenon suggests that competition is a motivating factor for bloggers to update and improve their articles.

\[\text{Figure 1. Sharing for shared understanding (item 1–4) and active learning (item 5)}\]

**Beyond sharing: Personal integration for active learning**

As Suthers (2005) suggests, the online replication of face-to-face learning is not acceptable as a CSCL goal; the same is true for using CSCL to duplicate social interactions over the Internet. Instead, educators should aim at using the unique features of the Internet as a large resource pool, especially its distribution characteristic (Scardamalia &
When designing BeyondShare, we purposefully implemented the sharing construction principle (Resnick, 1996) to encourage students to share and reuse ideas from each other’s constructions. Examples of approaches that require students to reuse or model parts of their peers’ projects to enhance their own personal integration include LEGO Mindstorms (Resnick, 2002) and Knowledge Soup (Canas et al., 2001).

In addition to shared constructions, we injected a sense of competition into BeyondShare to promote active engagement. As depicted in item 5 of Figure 1, students become active learners for the purpose of integrating personal knowledge. They are encouraged to evaluate their peers’ efforts regarding other learning units, select “personal best-fits,” and incorporate works they define as useful into their final personal products. Understanding of the learning material is strengthened through a process of incorporating ideas from their peers’ personal constructions as well as through reflecting on feedback concerning their own constructions.

Students compete to have their constructions selected by others as the most useful. As with bloggers, competition is used to motivate students to create, update, and upgrade quality products to share with others, as well as to evaluate their peers’ work in a serious manner. Through this competition, they gain a more comprehensive understanding of the learning material. Each student plays several roles and has specific responsibilities throughout an activity. The interchangeability of those roles encourages students to become active learners rather than passive information receivers (Table 1). Details will be described in the Procedure subsection of the BeyondShare Evaluation section.

Table 1. Beyond-sharing activity structure

<table>
<thead>
<tr>
<th>Expected learning outcome</th>
<th>Task unit</th>
<th>Student role interchange</th>
<th>Cooperation goal</th>
<th>Learning format</th>
<th>BeyondShare support*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Construct a personal concept map</td>
<td>Within a given unit</td>
<td>Active sharer vs. passive recipient</td>
<td>Personal accountability</td>
<td>Meaningful learning: reading, understanding, organization</td>
<td>Personal construction interface</td>
</tr>
<tr>
<td>2. Compete to be chosen by other students</td>
<td>Within unit, cross-unit</td>
<td>Within-unit competitor vs. cross-unit helper</td>
<td>Positive task interdependence via sharing cross-unit concept maps; sense of competition enhance motivation</td>
<td>Social facilitation and modeling</td>
<td>Personal construction interface; sharing construction interface</td>
</tr>
<tr>
<td>3. Evaluate and compare peers’ concept maps</td>
<td>Cross-unit</td>
<td>Peer assessor vs. receiver of peer feedback</td>
<td>Help peers revise their work; gain information about other units</td>
<td>Active learning: critical evaluation</td>
<td>Sharing construction interface</td>
</tr>
<tr>
<td>4. Construct an integrated concept map</td>
<td>Based on a given unit to link across all units</td>
<td>Active integrator</td>
<td>Based on a given unit for interlinking concepts across all units</td>
<td>Active learning: integrate personal and peers’ ideas according to a meta-plan</td>
<td>Sharing construction interface</td>
</tr>
</tbody>
</table>

*Note: See “Primary Interface” section

The term “beyond sharing” refers to combining the features of structuring and competition to achieve the goals listed in Table 1. Many new teachers initially assume that all learning (including listening to lectures) is inherently active. But the preponderance of research over the past few decades suggests that students must do more than just listen — they must actively discover and understand facts through reading and discussion, then transform and construct knowledge by writing or engaging in problem solving (Johnson, Johnson, & Smith, 1998; Moreno & Mayer, 2000). Active involvement means that students must engage in higher-order thinking tasks that entail analysis, synthesis, and evaluation (Turner et al., 1998). BeyondShare promotes active learning by encouraging: a) deep understanding of learning material via concept map construction (what Novak & Gowin [1984] refer to as “meaningful learning”); b) active reflection on the quality of individual constructions through sharing and peer evaluation; and c) the active synthesis of dispersed knowledge by integrating self- and peer-produced constructions (Fig. 4).
Peer assessment

Peer assessment is a widely used strategy in secondary and post-secondary classrooms for teaching principles in such diverse fields as writing, teaching, business, science, engineering, and medicine (Falchikov, 1995; Freeman, 1995; Rada, 1998; Strachan & Wilcox, 1996). The process requires such cognitive activities as reviewing, summarizing, clarifying, giving feedback, diagnosing errors, and identifying missing knowledge or deviations from an ideal (Van Lehn, Chi, Baggett, & Murray, 1995). Receiving abundant and immediate feedback from peers is strongly correlated with effective learning outcomes (Bangert-Drowns, Kulick, Kulick, & Morgan, 1991; Crooks, 1988; Kulik & Kulik, 1988). In conventional classroom settings, teacher feedback may be of higher quality but less frequent and immediate than peer assessments (Topping, 1998). In peer-assessment scenarios, students have more opportunities to view a larger number of projects, allowing them to gain inspiration from concrete examples instead of relying on models centered on a teacher's cognitive skills or knowledge structure. Peer-assessment projects require more on-task time than conventional teacher assessment settings; arguably this is the most important factor in facilitating learning.

Falchikov & Magin (1997), Lin, Liu, & Yuan (2002), and Liu, Lin, & Yuan (2002) are among researchers who state that reliable and valid peer assessment requires three conditions: a) students must fully understand and be committed to the purpose of their assessment activities; b) students need to be involved in the process of determining criteria, rating scales, and assessment procedures; and c) students need to receive feedback on peer-assessment scores in relation to their own performance as well as to the overall score pattern.

The BeyondShare environment

We incorporated concept mapping into the BeyondShare environment as an activity based on the assertions of Novak and Gowin (1984), Roth and Roychoudhury (1992), and others that concept maps are effective tools for knowledge construction. Instead of requiring students to participate in group discussions to create collaborative maps (a process that can lead to unequal contributions), we applied the CCL strategy (Lin, Sun et al., 2002) as a more effective approach to evaluating, synthesizing, and incorporating ideas from maps created by their peers. In implementing this strategy, the learning material must be divided into several units (in this study, three units). As part of the BeyondShare process, final concept map products reflect individual and shared construction efforts that fulfill the requirements of independence and interaction (Katz, 2002). In classrooms that have access to state-of-the-art learning technologies, teachers can use concept map approaches that focus on synchronous (real-time) cooperative behavior (Komis, Avouris, & Fidas, 2002). Although these systems have clear advantages, we purposefully designed BeyondShare with the characteristic of asynchronous distributed learning that we believe is available in a larger percentage of classroom settings.

Primary interfaces

We used a combination of Microsoft Visual Basic 6.0 and SQL Server7 to design two BeyondShare interfaces:

1. A personal construction interface that provides a form-based environment. This interface is disabled when students proceed to the sharing construction phase, thereby preventing students from modifying their own concept maps based on the work of others in the same learning unit (Fig. 2). After reading personal assignments for a given learning unit, students begin the personal construction activity in the concept mapping section by pressing the start button (which triggers a time log) and using the construction forms to build and connect self-defined concept nodes with links. A concept map in progress is shown in the current personal concept map section. Concept nodes and linking words are not fixed, giving students greater flexibility for knowledge construction. They use the current personal concept map area to observe and change node positions to revise concept hierarchies. Nodes and linking words can be removed from the storage section once they become irrelevant to the concept map. Students move back and forth between procedures to construct their maps as they see fit.

2. A form-based sharing construction interface consists of interlinks among different concept maps. Interlinks differ from links, which connect ideas within individual concept maps. In Figure 3, the bold arrows with dashed lines indicate interlinked connections between two concept maps. Students can use this interface to view their
own completed maps in the personal concept map section. In the modeling section, a system of anonymous selector IDs prevents students from purposefully choosing concept maps made by their friends as favorites. After choosing selector IDs from the other units, students can study maps in their peer concept map sections, then press the start button to begin the sharing construction process. Students can establish interlinks between their own and their chosen maps in the interlinking section and make comments in the feedback section according to a set of reference criteria. As in the personal construction interface, students can delete interlinks displayed in the storage section. The interlinking process consists of selecting single concept nodes from two maps and adding a linking word. Students can establish as many interlinks as they want between concept map nodes.

Figure 2. Example of personal construction interface

![Personal Construction Interface](image)

Figure 3. Example of sharing construction interface

![Sharing Construction Interface](image)
During the sharing construction phase, students evaluate all peer concept maps in other units, select “personal best fit” concept maps, and establish interlinks between their own and selected maps. Interlinks can be established between near concept nodes or nodes in remote categories. Links in the latter category are known as “cross-links,” implying associations between concepts that many people would not recognize (Novak & Gowin, 1984). In BeyondShare, such links are considered signs of creativity.

Choices for establishing interlinks represent cooperative partner selection — the result of a peer-assessment evaluation process that encourages critical thinking. Sharing and incorporating information across units with cooperative partners are both encouraged; within units, competition is encouraged.

**Teacher observation**

BeyondShare contains a teacher interface for monitoring student progress, meaning that students who fall behind the learning schedule can be given special attention. The monitor interface presents a student’s personal concept map, information on the student’s chosen favorites, the number of interlinks between two maps, how much time a student spends on constructing interlinks, and how many other students choose the same map as their favorite. The interface also allows teachers to view information on how many choose the target student’s concept map as their favorite, their personal concept maps, and respective interlinks. All preference data can be logged for peer rating analysis.

**Evaluating results**

After the sharing construction phase is completed, concept maps are arranged in decreasing order of score (number of votes) for each learning unit. The map receiving the most votes within one unit earns the designation of “best fit.” Reflective thinking is triggered via comparisons of personal maps with best-fit maps. Furthermore, teachers can construct their own “expert” concept maps for comparison with best-fit maps for two purposes: determining which knowledge structures are acknowledged by the greatest number of students, and helping students make adjustments to incomplete or incorrect concept maps.

**BeyondShare evaluation**

**Participants**

A BeyondShare evaluation test was conducted to determine if the beyond-sharing activities successfully engaged students in meaningful learning and knowledge construction. Participants were 34 college freshmen enrolled in an introductory computer science class at a research university in northern Taiwan. Students were randomly assigned to three clusters consisting of 12, 9, and 13 participants, with students in each cluster studying one of three learning units on the topics of function, class, or flow as selected from a C++ textbook. Members of each cluster generated individual concept maps for their assigned unit.

**Procedure**

We purposefully designed a series of beyond-sharing activities to ensure active learning, positive interdependence, and personal accountability. Using BeyondShare features, cooperative learning was structured by having participants work on a multiple-stage concept-mapping task requiring task interdependence (Table 1). After being grouped according to learning material divisions, students were asked to produce their own concept maps (a task that Novak & Gowin [1984] refer to as “meaningful learning”) for their assigned unit and to share their products with peers who worked on other units. Participants were instructed to evaluate, compare, and give feedback for the cross-unit concept maps. Participants therefore contributed to their classmates’ tasks by giving feedback while gaining information and knowledge about the other learning units. Based on a meta-plan, participants were asked to link their own maps with the cross-unit maps they selected during the peer-assessment stage to form integrated maps. Participants accepted responsibility for contributing to their cross-unit peers’ efforts while competing with same-unit peers. Participant roles switched among active and passive sharers, competitors and helpers, assessors and feedback
As shown in Figure 4, the evaluation procedure consisted of three stages:

1. Preparation. During week eight of the school semester, students were taught concept-mapping techniques and given several examples for practice. During week nine they were introduced to BeyondShare and its activities, after which they were randomly assigned to one of the three units.

2. Personal construction. During week ten, participants used their class time to create their individual maps. In an attempt to prevent social loafing or duplications of their classmates’ efforts, the students were not allowed to view their peers’ maps during this stage.

3. Sharing construction. During week eleven, students were allowed to view the concept maps created by classmates assigned to the other units. They were instructed to select one personal best-fit map from each unit and to establish interlinks across units. Participants were explicitly instructed to make their selections in terms of cohesiveness and coherence, and to avoid making their selections based on friendship or exchanges of favors.

At the end of week eleven, students were asked to complete a questionnaire about the BeyondShare environment and their subjective experiences with and perceptions of the beyond-sharing activities.

**Scoring**

1. Personal construction peer rating. Concept maps could be selected by peers assigned to other units based on general appearance or a specific task perspective (e.g., the best fit with a student’s own work). The number of votes thus represents the degree of cohesiveness and/or coherence between the concepts and structure of two maps. Personal construction scores accounted for 60% of the total peer rating, reflecting our goal of emphasizing personal accountability in active learning.

2. Sharing construction peer rating. Based on evaluations of cohesiveness and coherence, this rating (which accounts for 40% of the peer rating total) represents the number of votes earned by an individual student’s favorite maps.

   \[
   \text{Total peer rating} = 0.6 \times \text{personal construction}_{\text{peer}} + 0.4 \times \text{sharing construction}_{\text{peer}}.
   \]

The proposed peer rating system mimics the system of scholarly journal citations — that is, the more citations (votes) a work gets, the more likely the chosen work is of high quality. However, BeyondShare also takes into account the quality of the selected works. In other words, students must take responsibility for their personal best-fit choices because the scores of their selected maps affect their own final scores. This mechanism reduces the odds of students choosing maps created by their close friends regardless of quality.
Questionnaire

A questionnaire was created to measure the participants’ subjective perceptions of BeyondShare and beyond-sharing activities. The first section consisted of six items on interface usability, such as clarity of screen design, function simplicity and helpfulness, and comparative convenience.

Both time-spent and screen-capture records of construction procedures during the personal and sharing construction stages can serve as measures of active learning. However, it is important to note that active learning can take the form of a few meaningful and effective construction steps being produced quickly, or carefully planned cognitive functions emerging over a long time period. We therefore relied on a combination of learning outcomes and questionnaire responses to estimate how many participants felt that they were engaged in active learning and to gather supporting evidence for their responses.

The nine items in the second section focused on student perceptions regarding personal construction (first-level beyond-sharing activity) and approaches to active engagement in meaningful learning. The next six items measured if and how peer assessment (second-level) and competition influenced active engagement in knowledge construction. The final six items recorded student perceptions on sharing construction (third-level) and approaches to knowledge sharing. Responses were measured along a seven point Likert-type scale, with 1 indicating strong disagreement and 7 indicating strong agreement.

Results and discussion

All 34 participants had sufficient time to finish their personal construction projects and to evaluate, select, and integrate ideas from their peers’ maps into new, integrated concept maps. Participants needed an average of 2.15 hours to construct their personal concept maps following two one-hour introductions to concept mapping and BeyondShare. Average time spent in the sharing construction process was 1.07 hours. Sample concept maps are shown in Figure 3. Just over one half (53.5%) of the participants reported positive attitudes about the general ease of use of BeyondShare (Table 2). (In this and other tables, boldface indicates data that have been added together and presented as a sum.)

Concept mapping has been criticized for requiring exceptional effort and numerous modifications (Ruiz-Primo & Shavelson, 1996). The questionnaire data indicate that 39.3 percent of the participants regarded concept mapping using BeyondShare as more convenient than using pencil and paper. Negative opinions regarding the procedure were reported by 28.6 percent — an indication that BeyondShare requires revision. Just over two thirds (68%) stated that the personal construction interface was helpful, with 71.4 percent describing the interlink function as easy to use.

<table>
<thead>
<tr>
<th>Table 2. Student perceptions of NetShare’s ease-of-use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of respondents</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>1. In general, NetShare was difficult to use.</td>
</tr>
<tr>
<td>2. Creating a concept map using NetShare was more convenient than using pencil and paper.</td>
</tr>
<tr>
<td>3. The personal construction interface was clear and its functional guides were helpful.</td>
</tr>
<tr>
<td>4. The personal-construction visual aids were helpful when creating a concept map.</td>
</tr>
<tr>
<td>5. The interlink function procedure was simple and thoughtfully designed.</td>
</tr>
<tr>
<td>6. The system operating description was helpful when I first became acquainted with NetShare.</td>
</tr>
</tbody>
</table>
Table 3. Student perceptions of personal map constructions (first level)

<table>
<thead>
<tr>
<th>Percentage of respondents</th>
<th>strongly disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Constructing a concept map helped me in memorization.</td>
<td>7.1 57.1 17.9</td>
<td>0.0 3.6 3.6 10.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. When constructing a personal concept map, I had a chance to summarize critical points of the material.</td>
<td>42.9 28.6</td>
<td>14.3 39.3 32.1</td>
<td>0.0 3.6 3.6 7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I tried to use examples of events or concepts outside of textbooks to clarify the meaning of my concept map.</td>
<td>25.0 17.9 7.1</td>
<td>3.6 7.1 21.4 17.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Constructing a concept map encouraged me to rethink relationships between concepts.</td>
<td>21.4 25.0 32.1</td>
<td>0.0 7.1 0.0 3.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Constructing a concept map helped me organize key points in the learning material.</td>
<td>17.9 28.6</td>
<td>3.6 0.0 3.6 14.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. When constructing a personal concept map, organizing a concept hierarchy encouraged me to rethink knowledge synthesis.</td>
<td>17.9 50.0 17.9</td>
<td>3.6 0.0 3.6 7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. When constructing a personal concept map, I understood some of my shortcomings regarding the learning concepts.</td>
<td>0.0 3.6 7.1 14.3</td>
<td>14.3 35.7 17.9 25.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Although concept mapping was beneficial for meaningful learning, I felt it was not worth the trouble.</td>
<td>0.0 3.6 3.6</td>
<td>0.0 3.6 7.1 14.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I am willing to construct concept maps to aid my learning in other courses.</td>
<td>0.0 3.6 7.1 14.3</td>
<td>17.9 28.6 28.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 3, large percentages of students (50–89%) reported that they had actively engaged in the following cognitive functions:

1. Memorization (item 1): 82.1 percent agreed with the statement that concept map construction is an effective way to memorize learning material.
2. Summarization (item 2): 85.7 percent agreed with the statement that concept map construction gave them opportunities to summarize the most important points of the presented material.
3. Understanding (item 3): 50 percent stated that they used other materials in addition to textbooks when searching for examples that would give them a deeper understanding of a concept.
4. Conceptual organization (items 4, 5, 6): 89.4 percent asserted that drawing a concept map enhanced their comprehension of relationships between concepts, 85.7 percent stated that constructing a concept map helped them organize major concepts, and 78.5 percent agreed that concept hierarchy organization encouraged knowledge synthesis.
5. Reflections on own weaknesses (item 7): 85.8 percent agreed with the statement that drawing a concept map helped them reflect on their deficits, discrepancies, and/or flaws in learning concepts.

Only 7.2 percent of the participating students stated that concept map construction was not helpful in the learning process (item 8). The majority (75.1%) stated a willingness to construct concept maps to facilitate learning in other courses (item 9). We therefore suggest that personal construction (first-level beyond-sharing activity) encouraged student engagement in low- and high-level cognitive strategies and meaningful learning. This result fits well with the active learning and higher-order thinking criteria described by Johnson et al. (1998), Moreno and Mayer (2000), and Turner et al. (1998).

Data on responses to peer-assessment and competition (second-level) items are shown in Table 4. A majority (82.2%) agreed that the peer-assessment procedure helped them learn how to assess concept map quality (item 1) and 82.1 percent agreed that peer concept-map evaluation encouraged them to reflect on properties that a good concept map should possess (item 2).

Most of the participants (75.1%) stated that they were aware of the competitive aspect of BeyondShare and viewed it as motivation to generate better personal construction products (item 3); 74 percent acknowledged that they were
expected to compete with their peers for best-fit map votes (item 4). According to these results, the majority of participants were motivated to achieve personal learning goals when constructing quality maps. We suggest that this awareness of competition can reduce social loafing during beyond-sharing activities.

Approximately one-fifth of the participants (18.5%) complained about their maps not receiving votes even though they felt the quality was high (item 5), and 30 percent complained about a lack of satisfaction with their choices (i.e., they felt forced to choose from collections of poorly constructed maps) (item 6). A discussion mechanism such as that integrated by Scardamalia and Bereiter (1991) into their CSILE might help resolve this issue by encouraging modifications that increase map quality and/or coherence.

Table 4. Student perceptions of peer assessment and competition (second level)

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5  6  7</td>
</tr>
<tr>
<td>1. I learned how to assess concept map quality by evaluating and choosing concept maps from other learning units.</td>
<td>0.0 3.6 3.6 7.1 17.9 46.4 17.9</td>
</tr>
<tr>
<td>2. Evaluating and choosing a concept map encouraged me to consider the essential features of a “good” map.</td>
<td>0.0 3.6 3.6 7.1 21.4 28.6 32.1</td>
</tr>
<tr>
<td>3. Competition with peers to have my map selected as “best fit” for my unit encouraged me to generate a better personal construction.</td>
<td>0.0 7.1 3.6 10.7 17.9 42.9 14.3</td>
</tr>
<tr>
<td>4. I tried to gain more votes for “best-fit” concept map by generating a better personal construction.</td>
<td>0.0 7.4 3.7 14.8 25.9 33.3 14.8</td>
</tr>
<tr>
<td>5. I felt that the work I did was good, yet my peers did not chose my map as their favorite.</td>
<td>7.4 7.4 14.8 44.4 11.1 7.4 0.0</td>
</tr>
<tr>
<td>6. During the interlinking stage, I felt dissatisfied with what I chose as my favorite concept maps.</td>
<td>0.0 14.8 22.2 29.6 29.6 0.0 0.0</td>
</tr>
</tbody>
</table>

Data on the extent to which sharing-construction (third-level) activities helped students achieve active learning using high- and low-level cognitive strategies are presented in Table 5. As shown, the majority (85.7%) viewed the sharing construction activity as an effective means of helping them inspect and model their peers’ maps (item 1); 78.6 percent stated that observing their peers’ concept maps helped them make improvements to their own (item 2). Over half (57.1%) acknowledged that the sharing process allowed them to summarize key concepts in the chapters they did not work on and therefore gain general knowledge of all learning units (item 3), 57.2 percent agreed with the statement that they had achieved an in-depth understanding of the target material via the sharing construction procedure (item 4), and 64.2 percent agreed that the sharing construction approach was meaningful because it provided opportunities to integrate concepts from different units (item 5). However, 78.6 percent agreed with the statement that it required much effort to create meaningful interlinks between concepts (item 6). In summary, between 57.1 and 85.7 percent of the participating students agreed that the BeyondShare approach encouraged them to use the cognitive functions emphasized by Johnson et al. (1998), Moreno and Mayer (2000), Novak and Gowin (1984), and Turner et al. (1998).

The actively engaged students created high quality concept maps for sharing, offered valid ratings of their peers’ concept maps, and constructed coherent global concept maps that integrated ideas from other units. Different combinations of high and low personal and sharing construction scores were used to create the four cells presented in Table 6. High scores indicate that the student’s work exceeded the mean. According to the peer rating scores, 38 percent were high active learners (i.e., active in both sharing and personal construction), 29% were active only in terms of sharing construction, and 9% were active only in terms of personal construction. In other words, approximately 75 percent were active in at least one part of the beyond-sharing activities and 25 percent were not active during any part of the BeyondShare evaluation project. According to these results, it was easier for the participating students to actively engage in sharing construction than in personal construction.
Table 5. Student perceptions of sharing construction (third level)

<table>
<thead>
<tr>
<th></th>
<th>Percentage of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
</tr>
<tr>
<td>1. The sharing construction process allowed me to model my peers’ work.</td>
<td>0.0 3.6 0.0 7.1</td>
</tr>
<tr>
<td>2. The sharing construction process gave me chances to observe my peers’ work in a manner that helped my subsequent work.</td>
<td>0.0 7.1 3.6 7.1</td>
</tr>
<tr>
<td>3. The sharing construction allowed me to concentrate on my own work while referring to others’ concept maps for quick impressions of the other learning units.</td>
<td>0.0 7.1 17.9 14.3</td>
</tr>
<tr>
<td>4. The sharing construction process helped me achieve an in-depth understanding of the learning material.</td>
<td>0.0 7.1 7.1 25.0</td>
</tr>
<tr>
<td>5. The sharing construction process, which encouraged me to integrate concepts from different learning units, was a meaningful learning approach.</td>
<td>0.0 7.1 10.7 14.3</td>
</tr>
<tr>
<td>6. It was difficult to think of meaningful interlinks between two concepts.</td>
<td>0.0 0.0 0.0 17.9</td>
</tr>
</tbody>
</table>

Table 6. Scores on personal and sharing construction

<table>
<thead>
<tr>
<th></th>
<th>Sharing construction scores</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Personal construction scores</td>
<td>13/34 (38.24%) – A</td>
<td>3/34 (8.82%) – B</td>
</tr>
<tr>
<td></td>
<td>10/34 (29.41%) – C</td>
<td>8/34 (23.53%) – D</td>
</tr>
<tr>
<td>Total</td>
<td>23/34 (67.64%)</td>
<td>11/34 (32.35%)</td>
</tr>
</tbody>
</table>

Conclusion

After defining four types of sharing construction (basic, with notification, with feedback, and with interaction), we proposed a structured method we refer to as beyond sharing to encourage personal integration for active learning. Most current Web 2.0 applications support knowledge sharing or cooperation tools that fit in the four categories. These tools provide rich opportunities for users to experience sharing using various media (e.g., text, photos, music, video clips) or to co-write articles (e.g., Wikipedia). Researchers can utilize these applications as new platforms in order to observe how sharing activities or cooperation evolves in computer-supported cooperative work (CSCW) environments. To achieve the benefits of learning in a CSCW environment, we emphasized its active learning aspects over simple information-sharing activities (although users can benefit from shared knowledge) by adopting a CCL strategy for structuring learning activities (Lin, Sun et al., 2002). By having competition injected into a sharing activity, students are motivated to elaborate on their knowledge for deeper understanding.

We suggest that BeyondShare is capable of eliciting active knowledge contributions and empowering users to accumulate knowledge via social construction. Engaging students in active learning was a specific focus of our evaluation test, that is, determining to what extent participants perceived other student maps as information resources and used that information to develop a sense of a learning community via peer assessment. Results from a formal evaluation with 34 Taiwanese college freshmen support BeyondShare’s ease-of-use and ability to promote active learning. The same results also indicate that a) students who did not have advanced computer/Internet skills found BeyondShare easy to use; b) the personal construction process helped create a sense of meaningful learning in terms of both low-level (e.g., memorization and summarization) and high-level cognitive strategies (e.g., deep understanding, conceptual organization, and reflection); c) the sharing construction process helped create a sense of meaningful learning in terms of low-to-high level cognitive strategies; d) peer assessment helped foster active learning; e) BeyondShare’s competitive aspect was generally viewed as a motivating factor; and f) approximately 25
percent of the participants were not active at all during the BeyondShare evaluation experiment and 75 percent were active during at least one part.

One study limitation is that the sample was relatively small and limited, that is, all students were recruited from a single class at one university. Sampling bias and participant homogeneity could detract from the generalizability of the findings. Researchers may be interested in testing BeyondShare or similar online learning environments with students at different age levels and from a variety of schools, as well as in determining whether the beyond-sharing concept can be applied to tasks associated with skills development, such as programming, graphic design, and webpage design. Others may be interested in using personality inventories such as Big Five Personality Traits (Saulsman & Page, 2004) or 16 Personality Factors (Conn & Rieke, 1994) to identify successful and less successful learner characteristics for beyond-sharing activities.

To our knowledge, BeyondShare is the first learning product aimed at combining the features of structuring and competition, which distinguishes it as an environment that serves an active learning purpose instead of using the Internet to simply share information. BeyondShare also differs from other systems in that it tries to achieve active learning by accommodating cooperation and competition. In other words, students must decide how to use or incorporate parts of their peers’ ideas into their own work for a more comprehensive understanding of a topic. During this process of integrating their concept maps with others, students gain a deeper understanding of material across several learning units.

We suggest that teachers interested in using BeyondShare develop comprehensive plans, giving special consideration selecting authentic learning materials to introduce the social construction concept to students, dividing the material into independent but related subtopics, teaching concept map skills, and giving direct instruction on how to use the program. During the personal construction phase, teachers need to closely monitor their students to make sure they adhere to the principles of personal accountability and are not intimidated by competition. During the sharing construction phase, teachers need to encourage peer observation, critical evaluation, sharing, and unbiased peer ratings.

We believe that learning activities should be structured to create a balance between cooperation and competition in order to enhance motivation and learning performance (Johnson et al., 1981; Tauer & Harackiewicz, 2004), but we also acknowledge the difficulty of maintaining such a balance. Teachers may find that some of their students are more focused on competition, indicating a need to emphasize other beyond-sharing activities and benefits. Some teachers may be interested in creating a greater sense of cooperation by asking certain groups to discuss and reach a consensus in terms of interlinks, thereby encouraging the collective consideration of high-quality concept-map properties. In short, teachers are encouraged to experiment with the BeyondShare environment to make learning activities either more competitive or more cooperative. The activities are sufficiently flexible to accommodate these kinds of modifications.

References


Testing Principles of Language Learning in a Cyber Face-to-Face Environment

Nian-Shing Chen
Department of Information Management, National Sun Yat-sen University, Taiwan // Tel: +886 7 5252510 // Fax: +886 7 5254799 // nschen@faculty.nsysu.edu.tw

Yuping Wang
School of Languages and Linguistics, Griffith University, Australia // Tel: +61 7 3875 7537 // Fax: +61 7 3875 6766 // y.wang@gu.edu.au

ABSTRACT
This article discusses the application of the established principles of instructed language learning in a cyber face-to-face environment supported by an advanced Synchronous Learning Management System (SLMS). Following a critical review of the use of SLMS in distance language learning, the main body of the article focuses on the discussion of results from an empirical study, in reference to the relevant principles of instructed language learning formulated by Ellis (2005). These results indicate that principles of instructed language learning for traditional classroom-based learning are equally applicable in cyber face-to-face learning, but with their own characteristics. Recommendations for the effective use of SLMSs for online synchronous language learning are also put forward.

Keywords
Synchronous learning management system, Distance language learning, Cyber face-to-face interaction, Principles of instructed language learning, Synchronous cyber classroom

Introduction
In the last 20 years or so, research in CMC (Computer Mediated Communication) has witnessed tremendous developments both in terms of theory and practice. However, what is still needed, as White (2006: 249) identifies, “is the development of research tools, methods and approaches appropriate to the new paradigms for distance language learning”. There has been little research that has systematically tested principles of instructed language learning in a distance context supported by online Synchronous Learning Management Systems (SLMSs). Different from standalone technologies such as videoconferencing and audio conferencing tools, a SLMS employs a combination of synchronous and asynchronous technologies such as audio, video and other data sharing and interaction tools (e.g., an interactive whiteboard, document sharing, desktop sharing and multimedia discussion forums) to provide an integrated learning environment where multimodal interaction with both human and materials can be facilitated. Such interaction resembles most closely but is significantly different from face-to-face interaction in the traditional classroom. Thus the concept of “cyber face-to-face” interaction is used here to denote the differences. Since such a learning environment is still new to us, we are still unsure of its potentials for distance language learning which requires constant multi-way interaction in text, audio and video. For example, in what ways does cyber face-to-face interaction differ from face-to-face interaction in the traditional classroom? Can principles of instructed language learning be readily applied in a cyber classroom, and if so, in what ways? This article aims to answer these questions by testing some of the principles of instructed language learning in a SLMS-supported distance language learning environment. We will first review the current research development in the use of Learning Management Systems (LMSs) for facilitating interaction in distance language learning, and then present the ten instructed language learning principles proposed by Ellis (2005). The article will then discuss the results of an empirical study in which some of these principles were applied and tested. These results consist of examples of what happened in cyber face-to-face speaking tutorials, and will be discussed in reference to the relevant principles of instructed language learning, the characteristics of distance language learning and the potentials of cyber face-to-face learning. Recommendations for the effective use of SLMSs for online synchronous language learning will also be put forward.
The provision of interaction in distance language learning and the capacities of Learning Management Systems

Interaction, especially in the form of conversation, has long been considered crucial to L2 (Second Language) acquisition. However, the provision of interaction has also been the greatest challenge facing distance language learning. In other words, despite the availability of various educational technologies, learning to speak the target language still remains the most challenging of all the skills (e.g., reading and writing) that distance language learners must acquire. Hauck and Hampel (2005:259) claim that “providing opportunities for the development of interactive oral and aural skills has presented a major challenge for the course developer”. Until recently, the provision of oral and visual interaction in distance language learning has been technically unsatisfactory if not totally unavailable in some circumstances. As a result, many established principles for communicative and interactive L2 learning have not been tested in a rich media-supported environment.

To improve the provision of interaction in distance language education has long been the focus of many studies on CMC (e.g., Hastie et al., 2007; Kotter et al., 1999; Lamy, 2004; Levy & Kennedy, 2004; Shield et al., 2000; Sykes, 2005; Wang, 2004a, 2004b, 2006, 2007). CMC-based interaction can be categorized into three categories: text-based, audio-based and video-based (Wang 2004a). Many studies only concern themselves with the development of a certain skill (e.g. writing or speaking skill) supported by one type of the CMC tools (e.g., text chat, audio or videoconferencing). Research into the combination of text, oral and visual interaction supported by an integrated system such as a Learning Management System (LMS) is still lacking. The paucity of such research is partly due to the unavailability of effective LMSs for language learning, especially for synchronous distance language learning. Currently, LMSs can be roughly classified as Asynchronous LMSs (ALMS) and Synchronous LMSs (SLMS). The majority of the commercially available LMSs are asynchronous only, functioning as a teaching material management platform without the provision of synchronous text chat, data, audio and videoconferencing. Blackboard, WebCT, Moodle, and First Class (see Inglis, 2001) are some of the examples of an ALMS. Their functionalities include learning materials upload, discussion forums, individual or group emailing and assignment returns. Such systems are often used to supplement on-campus teaching. They are not ideal platforms for distance language learning as they do not support synchronous interaction in rich media forms such as cyber oral and visual interaction. As far as distance language learning is concerned, a SLMS is superior to an ALMS as, apart from the functionalities supported by an ALMS, it also offers synchronous conferencing supported by audio or videoconferencing tools, an electronic Whiteboard and other data conferencing tools. An ideal SLMS should support a learning environment in which multimodal interaction happens in real time in a synchronous manner. Such interaction, although different in many ways, resembles most closely physical face-to-face interaction, and can be used as a main platform for both managing and delivering distance language learning. For comparisons of LMSs, please see William Horton Consulting (2008), Web Conferencing Solutions (2008), and Alliance for Community Technology (2004).

The SLMS evaluated in this study is called Collaborative Cyber Community (hereafter 3C). The main reason for adopting 3C in our study was that it supports a much needed-dimension in distance language learning: quality cyber face-to-face interaction. 3C is both a learning management and delivery system facilitating the requirements of a complete language course offered online both asynchronously and synchronously. (For more discussion on 3C, see Chen, et al., 2005; Wang and Chen, 2007.)

We argue that it is time that the pedagogical values of the multimodal interaction supported by SLMSs was examined. This is more because of the urgent needs of distance learners for effective language acquisition than because of the availability and accessibility of advanced technologies. Thus the study presented here will investigate the ways in which principles of instructed language learning can be effectively applied in a cyber face-to-face environment supported by a SLMS.

Principles of instructed language learning

This study hypothesizes that established principles of instructed language learning for traditional classroom-based learning are equally applicable (but not totally replicable) in cyber face-to-face learning, despite the fact that there are significant differences between the two learning environments. This premise is grounded in our belief that the communicative nature of language learning in both environments remains the same and that technology is only a tool to help improve learning outcomes.
Ellis (2005: 209) draws together “findings from a range of second language acquisition studies” and formulates 10 general principles of instructed language learning. They are:

1. Instruction needs to ensure that learners develop both a rich repertoire of formulaic expressions and a rule-based competence.
2. Instruction needs to ensure that learners focus predominantly on meaning.
3. Instruction needs to ensure that learners also focus on form.
4. Instruction needs to be predominantly directed at developing implicit knowledge of the L2 while not neglecting explicit knowledge.
5. Instruction needs to take into account the learner’s built-in syllabus.
6. Successful instructed language learning requires extensive L2 input.
7. Successful instructed language learning also requires opportunities for output.
8. The opportunity to interact in the L2 is central to developing L2 proficiency.
9. Instruction needs to take account of individual differences in learners.
10. In assessing learner’s L2 proficiency it is important to examine free as well as controlled production.

These principles neatly sum up the current understanding of L2 acquisition theories and practices. Although Ellis (2005: 210) called these principles “provisional speculations”, they have been attested in L2 classroom research and practice. However, there has been little research on these principles applied in an cyber face-to-face environment supported by a SLMS, although some research has focused on one or two aspects of language learning (e.g., Hampel, 2006; Rosell-Aguilar 2005 on task design). To investigate in what ways some of these principles can be applied in cyber face-to-face learning is precisely the focus of our study. We will present findings from an empirical study in which some of these principles were applied in a distance language learning context supported by a SLMS, 3C.

![Figure 1. Major functions of 3C](image)

**Methods**

**The SLMS – 3C**

3C, the Collaborative Cyber Community learning platform, was developed by the National Sun Yat-sen University in Taiwan. Its server presently has a capacity to support up to 500 online asynchronous users and 200 online synchronous users simultaneously. Figure 1 summarizes the major functions of 3C.

As shown in Figure 1, 3C has two main environments: the “teacher’s office” and the “classroom”. The teacher’s office can be accessed only by the teacher, for student administration and planning learning activities, such as
uploading learning resources and designing content links. The “classroom” can be accessed by both the teacher and the learner, and it has two modes: the asynchronous and synchronous modes. The asynchronous mode is available to learners 24 hours a day, where audio, video and text-based learning resources (e.g., discussion boards, lecture notes, web-based course materials, and video recordings of cyber face-to-face sessions) can be accessed. As far as interactive language learning is concerned, the most valuable component in 3C is its synchronous mode with its main and sub cyber face-to-face classrooms. As shown in Figure 2, these classrooms feature five major windows: the main audio and video, the control panel, the text chat box, the Whiteboard and the sub-video windows. Up to 18 sub-video windows can be displayed at the same time. This cyber classroom is also supported by versatile synchronous data sharing tools, such as Desktop Sharing, Window Capture, Joint Web Browsing, Remote Control and collaborative annotation tools (e.g., pens and pen colours).

Figure 2. The synchronous cyber classroom

The procedure

A pilot study was carried out in late 2005 to investigate the application of some of the principles of instructed language learning in the 3C-supported environment. Seven (3 female and 4 male) adult students enrolled in the intermediate level of the Open Learning Chinese Program at Griffith University volunteered to participate in the
study. Five of them were from different parts of Australia, one from Hong Kong and one from the Czech Republic. Each participant was supplied with a Web camera, a headset and a handbook containing instructions on how to set up the web camera, headphones and the videoconferencing tool. Participants were asked to install the equipment and software by themselves and test their set up with a technician before the cyber face-to-face tutorials started.

A two-hour speaking tutorial each week for 10 weeks was conducted with these participants in the “cyber face-to-face classroom”. Both the teacher and the students attended the tutorials from home. In these tutorials, various meaning-based tasks (role plays, surveys, bingo games etc.) were guided by the teacher and performed by the participants. The audio and video quality of the tutorials was found to be satisfactory for language learning, even with students using a dial-up connection. Occasional video or audio delay was experienced but was not significant enough to affect learning outcomes (see, Wang and Chen, 2007). All the tutorials were digitally video-recorded by 3C.

Data analysis

The recordings of these tutorials were transcribed, capturing the interaction through text chat, audio, video and other online tools such as the Whiteboard and Joint Web Browsing. Segments of transcripts were selected for analysis in this article in accordance with their relevancy to the principles of language learning discussed here. The selected transcripts will be analysed using the Conversation Analysis methods as described by scholars such as Sacks (1992), Psathas (1995) and Hutchby and Wooffitt (1998). Conversation Analysis is a methodology used to analyse units or structures of a conversation, such as turn taking, in order to determine how interlocutors negotiate for meaning. To assist the analysis, screen activities of the tutorials were also captured to demonstrate what actually happened in the classroom.

We acknowledge that the number of participants is small. However, our analysis focuses on transcribed excerpts from the cyber face-to-face sessions, not on the number of replies from the participants. No post-test evaluation was deemed necessary to verify the pedagogical framework discussed here as one of the authors was the teacher who conducted the cyber face-to-face sessions.

Results

Since the research focuses on interaction in distance language learning, only the most relevant principles, namely, principles 2, 3, 6, 7, and 8, are evaluated here. These principles will be further discussed together with the data from this research.

Principle 2: Instruction needs to ensure that learners focus predominantly on meaning

A task-based approach was adopted, in order to create opportunities for learners to attend to the meaning of the language, i.e. “the highly contextualized meanings that arise in acts of communication” (Ellis, 2005: 211). In our cyber face-to-face sessions, focus on meaning was predominantly achieved during the completion of various meaning-based tasks. These tasks include group presentations, noughts and crosses listening exercises, discussing family photos, bingo games, surveys and role plays. The target language, Chinese, was used mostly in task completion. Figure 3 illustrates what happened in a bingo game on practicing numbers and prices in Chinese. Three of students compete by ticking the correct numbers in the game table while listening to the teacher calling out the numbers in Chinese.

We can see in Figure 3 that S1 (using a blue colour pen) wins the game because she is the first one to get three correct numbers in a row. S2 (red) is close behind and S3 (green) only gets two correct numbers. Other students are also involved in the game by typing the numbers in the text-chat box as they hear the prices being called (see the lower left corner of the screen in Figure 3).

Our data indicates that focus on meaning can be facilitated by many useful tools embedded in 3C. When designing a task, it is important to take into consideration the various tools offered by a SLMS. For example, the collaborative
annotation functions in 3C allow multiple participants to write, type, draw and edit at the same time. In such a learning environment, multiple sources of spontaneous input from a group of learners can occur, and group discussions, such as concept mapping and negotiation of meaning, can be effectively facilitated. The bingo game presented above represents a case in point. In this game, participants engaged in group activities using the collaborative annotation functions (i.e., the pen and pen colours) simultaneously. This example also shows that, although the bingo game was an activity for three students, other students could also participate at the same time through text chat.

Figure 3. A noughts and crosses game in Session 9

Principle 3 Instruction needs to ensure that learners also focus on form

Acknowledging the existence of various interpretations of focus on form, this research defines focus on form following Long (1991) and Long & Robinson (1998). Long (1991: 45-46) defines focus on form as an occasion that “overtly draws students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication”. Radwan (2005: 71) argues that “meaning-focused instruction alone may not be sufficient for learners to acquire the linguistic elements of the target language” and that there is “a positive relationship between drawing learner’s attention to the formal properties of the target language and language development”. This was also confirmed in our research. In this research, opportunities for focus on form were created on three levels: pre-task, during task and post task.

Focus on form prior to the start of task performance usually happened in the first 10 to 15 minutes of each tutorial when key grammatical points were explained to ensure their appropriate use in task performance. This process of focus on form was adequately supported by the various tools offered by 3C, such as the Whiteboard, text chat and the audio. For example, in Session 8, before asking the participants to conduct a survey about distance, the teacher uploaded a lecture note outlining the main structures to be used when talking about distance (see Figure 4) and
explained the structures through the audio, the video and the Whiteboard. She then checked the students’ understanding by asking them to make up sentences using these key structures (see Table 1). Figure 4 also shows that the key points on the lecture note were highlighted through colors and an animation pen.

Table 1. Session 8 - Talking about distance

<table>
<thead>
<tr>
<th>Time</th>
<th>Text chat</th>
<th>Oral interaction</th>
<th>Visual interaction</th>
<th>SLMS tools used</th>
</tr>
</thead>
<tbody>
<tr>
<td>43:32</td>
<td>T: How do you say ‘what place is your home close to’? Use the ǐ structure.</td>
<td>S2: Um…wojia li, ah, nijia li shenme difang haha… (Um, my home from, ah, which place is your home…)</td>
<td>T circles “what place is your home close to” on the lecture note.</td>
<td>Audio (via Mic/speaker)</td>
</tr>
<tr>
<td>43:38</td>
<td>S2: Um…wojia li, ah, nijia li shenme difang haha… (Um, my home from, ah, which place is your home…)</td>
<td>...</td>
<td>S2 places her hand on her mouth.</td>
<td>Visual (via web camera) The pointer for pointing to items on the lecture note.</td>
</tr>
<tr>
<td>44:01</td>
<td>S2: Ah…close to, um…wo wangle, duibuqi. (Ah, close to, um…I forgot. Sorry.)</td>
<td>T: This one, jin. S2: Ah, just ‘jin’.</td>
<td>S2 laughs.</td>
<td>The Whiteboard showing the lecture notes</td>
</tr>
<tr>
<td>44:07</td>
<td>T: This one, jin. S2: Ah, just ‘jin’</td>
<td>...</td>
<td>S2 shakes her head</td>
<td>Annotation pen for highlighting key points on the Whiteboard (see red circles)</td>
</tr>
<tr>
<td>44:15</td>
<td>S2: Aha, nijia li shenme difang jin ma? Ah, just ‘jin’. (Aha, which place is your home close to?)</td>
<td>S2 looks down and looks up again.</td>
<td>T laughs</td>
<td></td>
</tr>
<tr>
<td>44:16</td>
<td>T: Yeah. No ‘ma’, yeah.</td>
<td>T laughs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44:19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Lesson 8 - Focus on form prior to task performance
In the above excerpt, S2 was not sure about the use of “jin” in the beginning. After the teacher’s explanation, she was finally able to use the correct structure to render “which place is your home close to?” into Chinese. This type of focus on form was often controlled and guided by the teacher, with a specific linguistic aim to achieve.

In contrast, focus on form which happened during task completion was more random and spontaneous. This type of focus on form occurred when interaction in the target language broke down due to unfamiliar or unknown linguistic forms or due to teacher’s correction of grammatical mistakes made by the students. Interaction resumed when the breakdowns were repaired, usually through explanations or interactional modifications (see Table 2).

### Table 2. Session 7 - Focus on form during a role play

<table>
<thead>
<tr>
<th>Time</th>
<th>Text chat</th>
<th>Oral interaction</th>
<th>Visual interaction</th>
<th>SLMS tools used</th>
</tr>
</thead>
<tbody>
<tr>
<td>57: 49</td>
<td>S2: … ah, Yingwen bushi ni de muyu ma? (Right, I think studying English is very difficult. Um, isn’t…ah, isn’t English your mother tongue?)</td>
<td></td>
<td>S2 looks down and up again.</td>
<td>Audio (via Mic/speaker)</td>
</tr>
<tr>
<td>57: 59</td>
<td>S1: Um, dui, wo… wo hen xingyun, wo hen xingyun, (Um, right. I…I’m very lucky. I’m very lucky.)</td>
<td></td>
<td>T laughs</td>
<td>Visual (via web camera) The Whiteboard for showing role play brief.</td>
</tr>
<tr>
<td>58: 11</td>
<td>um… Aren’t I?</td>
<td></td>
<td>S1 laughs</td>
<td>Annotation pen for highlighting key points on the Whiteboard</td>
</tr>
<tr>
<td>58: 13</td>
<td>T: Dui bu dui. (Aren’t I?)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58: 15</td>
<td>S2: Shi bu shi. (Isn’t it?)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58: 17</td>
<td>T: Oh, ‘dui bu dui’ or ‘shi bu shi’ dou keyi. (Oh, both ‘duibudui’ and shibushi’ are fine.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58: 19</td>
<td>Yeah, yeah. And S1, … ‘Is English your mother tongue’, if you say ‘Yingyu bu shi ni de muyu ma?’, that means ‘Isn’t English your mother tongue?’. The grammar is OK, but not the meaning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58: 38</td>
<td>But if you say ‘Is English your mother tongue?’, how do you say it?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58: 44</td>
<td>S2: Ah…ah, Yingwen shi nide muyu ma? (Ah, ah, is English your mother tongue?)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: occasions of focus on form are highlighted in bold.

In Table 2, focus on form occurred twice during this segment of the role play between S1 and S2. In the first instance, when the teacher noticed an error in S2’s sentence, “Isn’t English your native language”, she did not intervene until she heard S1 say “Aren’t I” in English. She then explained “Aren’t I” and “Isn’t English your native language” together to minimize the number of interruptions to the role play.

However, in this research, more focus on form was done by the teacher at the end of task completion. To ensure the smooth and continued flow of task performance, the teacher usually refrained from intervening in the process unless students requested advice from her. Instead, the teacher took notes of the grammatical points that needed to be further clarified, and discussed them immediately after the task was completed. Table 3 illustrates this process of focus on form.

In the above exchanges, the teacher did not correct the mistake made by S1 in ‘tamen xianzai you feichang Zhongguo cha” until the role play finished, for fear of interrupting the dialogue. Instead, she picked it up through corrective feedback upon the completion of role play.
As shown in the above examples, 3C was found to be very effective in supporting focus on form through its audio and video, Whiteboard and annotation tools. Opportunities for focus on form before, during and at the end of task performance happened as naturally and effectively as in physical face-to-face interaction. Our research indicates that the principle that instruction needs to ensure that learners also focus on form was fully realized in 3C-supported learning environment.

<table>
<thead>
<tr>
<th>Time</th>
<th>Text chat</th>
<th>Oral interaction</th>
<th>Visual interaction</th>
<th>SLMS tools used</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:04:27</td>
<td>S1: Yi bei kafei? Aha. Um...xianzai tamen you feichang Zhongguo cha, um, ni xiang bu xiang... um, ni xiang... Zhongguo cha ma? (A cup of coffee? Aha. Um...now they have very Chinese tea. Um, do you want...um, do you want...Chinese tea?)</td>
<td>T writes down sth S1 looks up S1 looks away from the camera and laughs</td>
<td>Audio (via Mic/speaker)</td>
<td></td>
</tr>
<tr>
<td>01:04:51</td>
<td>S2: Dui, wo hen xihuan chi Zhongguo cha. (Yes. I like eating Chinese tea very much.)</td>
<td>S2 looks at the camera S1 looks down</td>
<td>Visual (via web camera)</td>
<td>The Whiteboard for showing role play brief.</td>
</tr>
<tr>
<td>01:04:58</td>
<td>S1: Hao. (Ok.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:04:59</td>
<td>S2: Oh, he Zhongguo cha. (Oh, drink Chinese tea.)</td>
<td></td>
<td></td>
<td>Annotation pen for highlighting key points on the Whiteboard</td>
</tr>
</tbody>
</table>

We recommend that students and teachers should utilize the tools (e.g., the pointer, the different colour schemes, the pen or the typing function) provided by a learning system to assist focus on form. Furthermore, prepared lecture notes were found to be very effective in focus on form prior task performance.

**Principle 6: Successful instructed language learning requires extensive L2 input**

Lack of both naturalistic and instructed input is a major deficiency in traditional distance language learning. Learners’ limited exposure to the target language often comes from textbooks and taped audio and video materials. Thus, “comprehensive input” deemed essential to L2 acquisition by scholars such as Krashen (1981), has been limited. Ellis (2005:217) suggests two ways to expose students to more input:

1. Maximise use of the L2 inside the classroom, ideally by using the L2 as a means and object of instruction.
2. Create opportunities for students to receive input outside the classroom, by (a) making resources available and (b) learner-training in how to effectively use the resources.

Our data indicates that 3C can adequately support these two ways of exposing learners to more L2 input. In our cyber face-to-face sessions, using Chinese as a means and object of instruction was done similarly to that in a traditional classroom. Similar to traditional classroom teaching, listening comprehension exercise and dictation formed another kind of exposure to target language input.

Different from the traditional classroom, 3C also offers versatile tools to expose learners to more extensive L2 input. For example, the function “Joint Web Browsing” allowed the whole class to view a website together in both English and the target language, be it a text, audio or video-based web resource. Figure 7 is a screen capture of the whole class viewing a map of Australia on the web. The students and teacher discuss Australian cities in Chinese while looking at the map together.

3C was also used, in a number of ways, to create opportunities for students to receive input outside the classroom. For example, the teacher emailed the participants an audio file recorded using the online recording function inbuilt in the multimedia Discussion Forum in 3C and asked the participants to listen to the audio file and answer questions. This audio file and questions were then discussed in the ensuing cyber face-to-face session. Figure 6 illustrates the functions embedded in the Discussion Forum.
Figure 5. Session 6 – talking about cities in Australia

Figure 6. The multimedia Discussion Forum and its main features

- The attached sound file
- The Whiteboard attachment
- Audio Recording mechanism
The cyber face-to-face sessions were also recorded through 3C and the video recordings were placed in the asynchronous classroom (see Figure 7) for the students to review the sessions by themselves after class and to consolidate the input they received in class. This is a valuable source of input as language learning requires constant consolidation after class. These recordings can also assist students to catch up with the classes they have missed.

Principle 7: Successful instructed language learning also requires opportunities for output

Complementary to the principle of providing opportunities for input, providing opportunities for learner output represents another important dimension in effective L2 instruction. Central to this principle is the notion that language production itself is a process of acquisition both in terms of linguistic and pragmatic competence. However, in most traditional distance programs, opportunities for extensive output have been practically missing, as learners have been studying alone. Our research indicates that 3C could effectively provide output opportunities through videoconferencing and other data conferencing tools, such as the Whiteboard and the text chat function. It has also been confirmed in our research that effective learner output is best catered for in task-based learning. Our research placed an emphasis on task performance to ensure that students were provided with opportunities for spontaneous output. The spontaneous output contained in Table 4 is a case in point. In Session 5, after a brief focus on form, Student 1 was given four minutes for an oral presentation in Chinese about herself (see Figure 8). In her talk, she utilized the Whiteboard to present pictures of her family, friends and her dog. Upon completion of this prepared task, she was then given an opportunity to play the role of a teacher and asked her classmates to answer questions regarding the content of her presentation (see exchanges in Table 4).

We can see that the dialogue in Table 4 was spontaneous and natural. The teacher sometimes took part in the output but rarely corrected students’ grammatical errors for fear of interrupting the flow of their output. As exhibited in Table 4, even though S1 should not use ‘ma’ in ‘Shibushi ni ma’ and ‘Shibushi... airen ma’, the teacher did not
correct her. The fact that the students could see and talk to each other through videoconferencing helped to create a communicative atmosphere for output. Opportunities for output such as the one presented in Table 4 were provided in every cyber face-to-face session.

Table 4. Session 5 – question and answer session with S1

<table>
<thead>
<tr>
<th>Time</th>
<th>Text chat</th>
<th>Oral interaction</th>
<th>Visual interaction</th>
<th>SLMS tools used</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:51:48</td>
<td>S1: Di san ge, shui shi… Sugelan ren? (No.3, who is…from Scotland?)</td>
<td></td>
<td>S1 looks down and up</td>
<td>Audio (via Mic/speaker)</td>
</tr>
<tr>
<td>01:51:50</td>
<td></td>
<td>S2: Shui shi Sugelan ren? Dui bu qi, wo ting bu dong. (Who is from Scotland? Sorry, I don’t understand.)</td>
<td>S2 knitted her eyebrows and looks puzzled.</td>
<td>Visual (via web camera)</td>
</tr>
<tr>
<td>01:51:56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:52:05</td>
<td>S2: Ah, ha, um, zhe shi…um ni de baba. (Ah, ha, um, this is…your father.)</td>
<td></td>
<td>S2 realizes and laughs.</td>
<td></td>
</tr>
<tr>
<td>01:52:09</td>
<td>S1: Ah, bu dui. (Ah, no.)</td>
<td></td>
<td>S1 laughs</td>
<td></td>
</tr>
<tr>
<td>01:52:13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:52:15</td>
<td>S2: Mama, mama, mama, ni de mama. (Mother, mother, mother, your mother.)</td>
<td></td>
<td>S2 rolls her eyes</td>
<td></td>
</tr>
<tr>
<td>01:52:21</td>
<td>S1: Wo mama shi Aodaliya ren. Wo baba shi Helan ren. (My mum is Australian. My father is from Holland.)</td>
<td></td>
<td>S2 laughs and sits back.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T laughs</td>
<td>S1 tilts her head to the left and smiles</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S2 puts her hands on the back of her head.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 5. Session 9 – interaction between three participants in a Cafe**

<table>
<thead>
<tr>
<th>Time</th>
<th>Text chat</th>
<th>Oral interaction</th>
<th>Visual interaction</th>
<th>SLMS tools used</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:04:51</td>
<td>S2: Dui, wo hen xihuan chi Zhongguo cha. (Yes, I like eating Chinese tea very much.)</td>
<td>T takes notes</td>
<td>Audio (via Mic/speaker)</td>
<td></td>
</tr>
<tr>
<td>01:04:58</td>
<td>S1: Hao. (Ok.)</td>
<td></td>
<td></td>
<td>Visual (via web camera)</td>
</tr>
<tr>
<td>01:04:59</td>
<td>S2: Oh, he Zhongguo cha. (Oh, drink Chinese tea.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:05:22</td>
<td>S3: Ni yao chi shenme? (What do you want to eat?)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:05:36</td>
<td>S1: S2, wo xiang jintian wo bu chi dongxi, wo tai pang. (S2, I don’t think I will eat anything today. I am too fat)</td>
<td>T laughs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:06:18</td>
<td>S2: Yibei kafei, hai you yibe… Zhongguo cha. (One cup of coffee, and one cup of Chinese tea.)</td>
<td>All laugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:06:26</td>
<td>S3: Zhongguo cha, yige Zhongguo cha, yige kafei? Zhe shi um…liang, liangkuai um…Meigu de yuan. (Chinese tea, one Chinese tea. One coffee? This is um…two, two dollars…American dollars.)</td>
<td>S3 puts his fingers on his temples.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:06:48</td>
<td>T: Haha, Meiyuan.</td>
<td>T looks down and</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Principle 8: The opportunity to interact in the L2 is central to developing L2 proficiency**

The importance of interaction to L2 acquisition has long been established. Theorists often cite Long’s (1981, 1983) Interaction Hypothesis to summarize the current understanding of interaction in L2 acquisition. When discussing Interaction Hypothesis, Long and Robinson (1998: 22) emphasize that “[p]articularly important is the negotiation for meaning that occurs more or less predictably in certain interactions”. They continue that negotiation for meaning elicits interactional modifications, negative feedback, including recasts and noticing that “increase input comprehensibility without denying learners’ access to unknown L2 vocabulary and grammatical forms”. However, due to the lack of effective technology to bridge the distance between the learner and education provider, distance language learners have been deprived of synchronous oral and visual interaction until very recently. With the rapid developments and improvements in SLMSs, such as 3C, the provision of interaction similar to that of physical face-to-face interaction has become a reality. Our data indicates that 3C provides various options to create opportunities for learners to interact in the target language. Through a web camera, the students can interact with one another and with their teacher(s) both orally and visually. Table 5 presents an example of the kind of interaction facilitated by 3C.
S2: S1, wo jintian mei dai qian, meiyou, mei dai qian. (I haven’t brought any money with me today. …no, haven’t got any money.)
S1: Duibuqi, qing ni zai shuo yibian? (Sorry, could you please say it again?)
S2: Wo mei dai qian. (I didn’t bring any money.)
S: Oh …
S3: Meiyuan. (US dollars.)
S2: Ah, wo neng xiwan. (Ah, I can wash dishes.)
S3: Shenme? (What?)
S2: Xiwan. (Wash dishes.)
S3: Xiwan? (Wash dishes?)
S2: Wo mei dai qian. (I didn’t bring any money.)
S3: She can wash dishes, haha… ah… haha… Hao le, hao le. (Ok, Ok.)

*T laughs
*S1 looks up when hearing S2 calling her.
*S2 and S1 laugh
*S1 still laughing
*T and S3 laugh
*T covers her mouth with her hand and laughs
*S2 laughs.
*S2 shakes her head.
*S3 realizes and laughs
*All laugh

The task depicted in Table 5 only requires that a dialogue be created between three students role-playing a scene in a cafe. No specific details of the dialogue were stipulated by the teacher. Instead the students had to create the dialogue spontaneously as the scene developed. After ordering the coffee and tea, S2 says to the waiter (S3) that she has not got any money on her and offers to wash dishes to pay for the drinks. Interactional modification occurred when S2 realizes that she should have said ‘drinking Chinese tea’, instead of ‘eating Chinese tea’. S3 also modifies his Chinese term for US dollars after receiving the correction from the teacher. Recast also happened in the dialogue when S2 repeats what she has said at S1’s request. S1 finally realizes that S2 has said she did not have any money on her. Negotiation of meaning occurred again when S2 offers to wash dishes. S3 finally notices the link between washing dishes and having no money. This process of negotiation of meaning was also supplemented by visual interaction demonstrated through the video (see Column 4 in Table 5). For example, to emphasize that she has not got any money on her, S2 also shakes her head.

**Discussions and recommendations**

Data from this research indicates that an effective SLMS, such as 3C, can support communicative and interactive L2 learning. This section further discusses the above presented results and makes recommendations in terms of the use of the following three tools in an SLMS, which are particularly important to language learning:

- Text chat,
- Joint Web Browsing,
- Sub cyber face-to-face classrooms
Text Chat

If used effectively, text chat can cater for interaction in many important ways. It can be used to supplement and complement the audio and video when the audio and video quality becomes problematic. More importantly, the different kinds of interaction facilitated by text, audio and video channels cater for students of different learning styles and language proficiency. For example, the shy students can participate in learning by typing their answers or comments in the text chat box while listening to others’ verbal exchanges. The major points of a verbal exchange can also be summarized here by the teacher or an advanced student for reviewing what has been said orally. Furthermore, teachers can maximize the opportunity to better facilitate individual learning through text chat, answering specific questions raised by individual students while other students are performing a task orally. One might argue that using text chat concurrently with oral and visual interaction might create information overload for some students and teachers. However, today’s digital generation has been accustomed to receiving information from multiple sources simultaneously.

No doubt, a class protocol needs to be established to regulate the use of the text-chat function. For example, text chat should not be used for chatting about things irrelevant to the content of the particular session. In particular, no swear words should be posted there. For an advanced class, students can be asked to text chat only in the target language. To be more effective, certain icons can be created and used for indicating actions. For example, in our research, “papapapa” stands for applauses and the longer it gets, the warmer the applause it would indicate. A series of question marks (i.e., ?????) means “I have a question”.

Joint Web Browsing

Joint Web Browsing was found to be a valuable tool as it provides wider ranging and more authentic input than a traditional classroom can. It allows instant access to voluminous web resources in the target language. For example, the teacher and the learner can browse a website for teaching Chinese and use the audio files on the website (e.g., MP3 files) as a listening comprehension exercise. Or the whole class can watch authentic video segments on Youtube, and use it as a catalyst for discussions in the target language.

Sub cyber face-to-face classrooms

Ellis (2005: 220) believes that passing control of the discourse topic to the learner ensures interaction beneficial to acquisition, and that small group work is where “acquisition-rich discourse is more likely to ensue”. Our data indicates that small group activities were also effectively facilitated by the sub cyber face-to-face classrooms. For example, in Lesson 4, the students were divided into groups and sent to the sub cyber classrooms to collaborate in a conversation in Chinese. It is recommended that this function should only be used when students have grown familiar with the learning system. This is because they would have to leave the main cyber classroom and operate the sub classrooms all by themselves. It was found that the ideal number of groups was between 2 and 5. This allowed the teacher enough time to visit each sub classroom while learners are practising in their groups. These classrooms can also be used to provide opportunities for interaction after class as they can be set up by learners themselves at anytime. The learner can accomplish a specific task with one or a group of fellow students in these classrooms, and record their task performance for the teacher or other students to view.

Conclusion

This study has tested some of the principles of instructed language learning in a cyber face-to-face environment supported by 3C. Although largely descriptive and qualitative in nature, this study has depicted the dynamics of this environment and explored its pedagogical values for distance language learning. The results of this research indicate that the present generation SLMSs can adequately support the realization of those principles gleaned by Ellis (2005). However, this conclusion does not mean that these principles are readily applicable in a cyber face-to-face environment. Nor does it mean that the application of these principles replicates what happens in a traditional classroom. On the contrary, the process in which those principles were applied was pedagogically different in many significant ways between the two environments. Hampel (2006: 111) claims that “an easy (and cheap) transposition
of face-to-face tasks to a virtual environment is not possible; instead, we have to ensure that tasks are appropriate to
the medium used and that we develop tasks that take into account the affordances (i.e., the constraints and
possibilities for making meaning) of the modes available”. The versatile tools in 3C created an acquisition-rich
environment in which synchronous text, oral and visual interaction happened spontaneously, and complemented one
another. Comprehensive input and output with distinctive characteristics of cyber face-to-face learning were
adequately facilitated.

This research points to the great potential of an advanced SLMS for distance language education. The multiplicity of
such an environment needs to be further explored. The technologies are in place and will certainly improve as time
goes on. How we use them to effectively and creatively support L2 learning depends on the teacher’s and the
learner’s familiarity with the tools and their intuition in language teaching and learning, in addition to being guided
by established principles.

While testing established principles of language learning, this research also recognizes a need for the formation of
new principles addressing the characteristics of online learning. Such principles can include:
- instruction needs to ensure the use of various online tools to maximize opportunities for learners to interact
spontaneously in the target language;
- successful instruction needs familiarity with the affordance of an online environment;
- instruction needs to ensure that teachers and learners develop technical skills/competency for using online
learning tools effectively and creatively.

Further studies are urgently needed to formulate and test principles of synchronous online L2 learning. At the same
time, we argue that, in addressing the uniqueness of online learning, research into SLMS-supported learning needs to
be grounded on the established principles of language learning. No matter how advanced a technology is, it is still a
tool to support learning. In an emerging area of research such as SLMS, it is important to avoid throwing out the
baby with the bathwater.

Acknowledgement

This study was supported by the National Science Council, Taiwan, under grant NSC95-2520-S-110-001-MY2.

References


synchronous online environment. ReCALL, 18 (1), 105-121.

Hastie, M., Chen, N. S., & Kuo, Y.-H. (2007). Instructional design for best practice in the synchronous cyber

design and tutor role. In Holmberg, B. & Shelley, M. & White, C. (Eds.), Distance education and languages:
evolution and change, Clevedon: Multilingual Matters, 258-277.

Press.

Innovation in open & distance learning: successful development of online and web-based learning, London: Kogan
Page, 88-99.


Effects of Text, Audio, and Graphic Aids in Multimedia Instruction for Vocabulary Learning

Daesang Kim
Office of Academic Affairs, Saint Mary-of-the-Woods College, USA // Tel: +1 812 535 5273 // dkim@smwc.edu

David A. Gilman
Department of Curriculum, Instruction, and Media Technology, Indiana State University, USA // Tel: +1 812 237 2925 // dgilman1@isugw.indstate.edu

ABSTRACT
This study is an investigation of the use of multimedia components such as visual text, spoken text, and graphics in a Web-based self-instruction program to increase learners’ English vocabulary learning at Myungin Middle School in Seoul, South Korea. A total of 172 middle school students (14 years of age) in five classes participated in the study. Each individual was required to complete several testing instruments such as a pretest, posttest, retention test, and attitude inventory. Participants learned better when they received “visual text and added graphics” or “visual text, added spoken text, and added graphics” instruction. Although the added multimedia components required learners to spend more time on the instruction, the extra time was not significant. The results lead one to conclude that an effective way to improve learning of English vocabulary is to offer graphics that illustrate what the vocabulary means.

Keywords
Multimedia Learning, Web-based Self-Instruction, Admissible Probability Measurement Procedures, English as a Foreign Language Vocabulary Learning

Introduction
With computer technology, Web-based learning has become a common choice in education institutions (Bauer, 2002, p. 31). Furthermore, the variety of media such as text, graphics, audio, and video for delivering content has attracted many instructors and students to use the Internet for distance education (Ali, 2003). These multimedia components get and hold learners’ interest, which many researchers believe is important when teaching the video generation (Jonassen, 2000, p. 208). Visual text and graphics are some of the most popular tools in on-line learning. In many cases, graphics can be used to represent important information and are often used for supporting text (Newby, Stepich, Lehman, & Russell, 1996, p. 103). Using these techniques, the most widely used asynchronous on-line learning tool is courses primarily posted in visual text and static graphics (Liles, 2004).

English as a Foreign Language (EFL) learners often adopt various strategies to memorize vocabulary words. For instance, vocabulary learning is often used with strategies such as word lists or paired associations in which new words are presented with their translations. These strategies often require learners to memorize pair associations directly (Sun & Dong, 2004). Meara (1996) found that many researchers in vocabulary learning studies have explored various methods of vocabulary presentation and their corresponding effectiveness in retention. Some earlier studies found the following:
• There is a limit to the number of words that can be learned at one time (Crothers & Suppes, 1967, chap. 4).
• Rote repetition appears less efficient than using spaced recall and structured review (Atkinson, 1972).
• Silent repetition and silent writing are less effective than repeating the words aloud (Meara, 1996).

Brown and Payne (Hatch & Brown, 1995, p. 383) have identified five steps to vocabulary acquisition: (a) having sources for encountering new words; (b) getting a clear image, either visual or auditory or both, of the forms of the new words; (c) learning the meaning of the words; (d) making a strong memory connection between the forms and the meanings of the words; and (e) using the words.

Recently, a number of researchers have discussed the benefits of presenting information using multimedia components such as visual text, spoken text, graphics, and videos on language learning (Al-Seghayer, 2001; Chun & Plass, 1997; Duquette & Painchaud, 1996; Ehsani & Knodt, 1998). In their studies, information presented in text,
spoken words, graphics, and video formats can be integrated to create an authentic, attractive, and multi-sensory language context for EFL learners (Sun & Dong, 2004). Kost, Foss, and Lenzini (1999) found that EFL learners performed better on both production and recognition vocabulary tests when they were allowed to use a combination of visual text and graphics.

Designing pedagogically effective multimedia instruction in language learning based on theories has been an important issue (Chapelle, 1998; Hoven, 1999; Liu, Moore, Graham, & Lee, 2002; Watts, 1997). Mayer and Moreno (2002) focused on a cognitive theory of multimedia learning which combines dual coding theory (Paivio, 1986, chap. 4; Sadoski & Paivio, 2001, chap. 3), cognitive load theory (Sweller, Van Merrienboer, & Paas, 1998), and constructivist learning theory (Novak, 1998, chap. 3; Vygotsky, 1978, chap. 6). From dual coding theory they adopted the idea that verbal stimuli and nonverbal stimuli detected by our sensory systems are processed in different systems of the brain (verbal system and nonverbal system). From cognitive load theory they adopted the idea that “humans are limited in the amount of information that they can process in each channel at one time” (Mayer, 2001, p. 44). Sweller et al. explained that redundant memory load is caused by “the presentation format of instructions extraneous load” (Tabbers, Martens, & Merrienboer, 2004, p. 72). Mayer and Moreno (2002) finally concluded that “presenting too many elements to be processed in visual or verbal working can lead to overload” (p. 111). They also took the idea from constructivist learning theory that “meaningful learning occurs when learners actively select relevant information, organize it into coherent representations, and integrate it with other knowledge” (p. 111).

Mayer and Moreno (2002) found the following interesting results:

- Providing words with narration and animation helped learners’ performance more than words alone.
- Reducing the number of unneeded words and sounds helped learners’ performance.
- Providing words with narration helped learners’ performance more than on-screen text.
- Providing words as narration and animation helped learners’ performance more than narration, animation, and on-screen text.

Reducing the amount of on-screen text makes more area available for graphics and labeled illustrations, which are necessary tools for teaching certain types of concepts. Some studies indicated that including the visual text in the illustration and labeling the illustrations improved learning (Koroghlanian & Klein, 2004).

However, as Mayer (2001) states, “all multimedia messages are not equally effective” (p. 79). For example, Mayer concluded “Schnotz, Bannert, and Seufert (in press) reported situations in which some learners reduced the amount of attention they paid to text when pictures were added” (p. 79). Tabbers et al. (2004) concluded that replacing visual text with spoken text and added graphics to the visual text both do not easily generalize to non-laboratory settings.

By better understanding the effect of individual components of multimedia, language educators will be able to design effective instruction for EFL learners. This study is an investigation of the effect of multimedia components such as visual text, spoken text, and graphics on increasing learning or decreasing redundant memory load in English vocabulary learning.

**Method**

The primary objective of this research is to study the effects of six methods of instruction in a Web-based self-instruction program: visual text (Group A), visual text and added spoken text (Group B), visual text, and added graphics (Group C), visual text, added graphics, and added spoken text (Group D), reduced visual text and added spoken text (Group E), and reduced visual text, added graphics, and added spoken text (Group F). The researchers investigated the use of multimedia components such as visual text, spoken text, and graphics in a Web-based self-instruction program to increase English vocabulary learning at Myungin Middle School (MMS) in Seoul, South Korea.

Multiple choice questions are usually used to test a student’s ability to recall information, to interpret data or diagrams, and to analyze and evaluate material. In this study, Shuford’s Admissible Probability Measurement (APM) procedure was used to reduce guessing scores on the multiple choice tests (Gilman, 1967, p. 27). Shuford claimed that APM procedures operate scoring systems with a very unique property that guarantees that any student can maximize his expected score if and only if the student honestly reflects his or her degree of belief probabilities. The
The formula depends only on the probability assigned to the correct answer and not on probabilities assigned to the other incorrect alternatives (Gilman, 1967, p. 27). The score obtained from an expressed probability to a correct response \( r_k \) is expressed as a function \( g_k (r_k) \) such that

\[
g_k (r_k) = \begin{cases} 
1 + \log r_k & \text{for } 0.01 < r_k \leq 1 \\
-1 & \text{for } 0.00 \leq r_k \leq 0.01
\end{cases}
\]

The possible scores range from minus one to plus one and are fairly related to the practice of giving the student one point for each correct answer and subtracting one point for each incorrect answer in order to discourage guessing in multiple choice testing. In other words, a value of 1 indicates that the student’s response is correct for a question and a value of -1 indicates that the student’s response is incorrect for a question. For instance, a student whose responses are all correct on a 30-item multiple choice test will receive 30 points. A student whose responses are all wrong for a 30-item multiple choice test will receive -30 points for the test.

There were two forms of response for each multiple-choice question. The first was for the response to each question \( (a, b, c, \text{ or } d) \). The second required the learner to write a number from 0 to 100 to indicate how sure the student was that his or her response was correct. For instance, a student who was 100% sure that the response was correct would put “100” in the second space. If he or she was not completely sure, the second space should contain a smaller number. This number constitutes the degree of certainty score. From there, the admissible probability score is obtained by applying the formula listed above. For instance, if the student was 100% sure and his or her response was correct for a question, the admissible probability score would be 1 for the question; if the student was 0% sure that his or her response was correct for a question, the admissible probability score would be -1 for the question. If the student’s response was wrong for a question, the admissible probability score would be -1 for the question.

### Research Questions

The research questions for this study were:

1. What are the differences in original learning among students who are taught under the six methods of instruction conditions as measured by raw score, mean degree of certainty estimate, and an admissible probability scoring procedure?
2. What are the differences in time to complete instruction among students who are taught under the six methods of instruction?
3. What are the differences in students’ attitude toward instruction among students who are taught under the six methods of instruction conditions as measured by their scores on the attitude inventory?

### Participants

A total of 172 middle school students (14 years of age) in five classes participated in the study. The students had no previous experience with computer-assisted instruction in English vocabulary learning. All sampling procedures were random. All participants were separated into six groups, 43 students (Group A), 22 students (Group B), 34 students (Group C), 24 students (Group D), 24 students (Group E), and 25 students (Group F). The groups that used spoken text format (Group B, D, E, and F) were smaller than the non-spoken text groups (Group A and C) because many computers in the classroom did not have sound systems with headsets or speakers.

### Materials

The topic of the Web-based self-instruction was English vocabulary learning. The design of the English vocabulary instruction was based on the following criteria:

1. The items of English vocabulary were of appropriate difficulty level for Korean middle school students.
2. Graphics supported visual text.
3. Graphics were available for cueing meaning of the word from static or animated images.
4. Spoken text was used as narration to support visual text.
5. Reduced the amount of visual text on a screen left more area available for graphics and spoken text.
The illustration of the criterion 3, 4, and 5 are shown in Figure 1 with one example.

The items of English vocabulary to be learned by students on the Web-based self-instruction program were: 1) tether, 2) wither, 3) tumble, 4) separate, 5) gorge, 6) fetter, 7) beacon, 8) crest, 9) awl, 10) ditch, 11) entice, 12) taut, 13) quench, 14) wizen, and 15) waylay. The length of each lesson was a maximum of 30 minutes and the time was controlled by computer. The six groups of Web-based self-instruction program are shown in Figure 2.

**Procedures**

Each participant was required to complete a pretest, posttest, retention test, and attitude inventory for the study. A pretest was administered to the participants one week prior to the study. All students among the six different groups were required to respond to 30 questions regarding the vocabulary with a number from 0 to 100 to indicate their degree of certainty on the pretest. One week later, all students among the six different groups received multimedia instruction through a Web-based self-instruction program. Items of vocabulary were projected on the computer screen through the program. Then, all students among the six different groups were required to respond to 30
questions regarding the vocabulary with a number from 0 to 100 to indicate their degree of certainty on the posttest. Approximately one week later, all participants were required to again respond to 30 questions regarding the vocabulary with a number from 0 to 100 to indicate their degree of certainty on the retention test. In addition, all students were required to complete the 40 items of the attitude inventory.

Figure 2: An example showing frames of Groups (Translation into English)
Results

Data were analyzed using two analyses - mixed factorial design (the split-plot analysis of variance) and one-way ANOVA - to evaluate the results from the study with regard to the following variables:

1. Student’s raw scores on
   - Pretest
   - Posttest
   - Retention test

2. Student’s mean degree of certainty estimates on
   - Pretest
   - Posttest
   - Retention test

3. Shuford Admissible Probability Scores on
   - Pretest
   - Posttest
   - Retention test

4. Time required to complete multimedia instruction

5. Student attitude inventory

Table 1 presents the results of analyses of all variances with mean scores and standard deviation for each group on the pretest, posttest, and retention test.

<table>
<thead>
<tr>
<th></th>
<th>Group A (n = 43)</th>
<th>Group B (n = 22)</th>
<th>Group C (n = 34)</th>
<th>Group D (n = 24)</th>
<th>Group E (n = 24)</th>
<th>Group F (n = 25)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Raw Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>M = 10.67</td>
<td>M = 9.77</td>
<td>M = 9.65</td>
<td>M = 10.96</td>
<td>M = 10.71</td>
<td>M = 11.40</td>
</tr>
<tr>
<td></td>
<td>D = 5.43</td>
<td>D = 4.05</td>
<td>D = 3.01</td>
<td>D = 5.34</td>
<td>D = 3.45</td>
<td>D = 6.19</td>
</tr>
<tr>
<td>Posttest</td>
<td>M = 25.21</td>
<td>M = 21.95</td>
<td>M = 27.44</td>
<td>M = 26.17</td>
<td>M = 21.08</td>
<td>M = 25.48</td>
</tr>
<tr>
<td></td>
<td>D = 6.30</td>
<td>D = 7.31</td>
<td>D = 3.90</td>
<td>D = 6.19</td>
<td>D = 6.85</td>
<td>D = 6.83</td>
</tr>
<tr>
<td>Retention Test</td>
<td>M = 21.35</td>
<td>M = 17.82</td>
<td>M = 23.97</td>
<td>M = 25.42</td>
<td>M = 18.42</td>
<td>M = 23.08</td>
</tr>
<tr>
<td></td>
<td>D = 8.37</td>
<td>D = 9.53</td>
<td>D = 7.47</td>
<td>D = 4.85</td>
<td>D = 8.77</td>
<td>D = 8.95</td>
</tr>
<tr>
<td><strong>Degree of Certainty Estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>M = .32</td>
<td>M = .33</td>
<td>M = .32</td>
<td>M = .24</td>
<td>M = .21</td>
<td>M = .35</td>
</tr>
<tr>
<td></td>
<td>D = .29</td>
<td>D = .28</td>
<td>D = .28</td>
<td>D = .17</td>
<td>D = .21</td>
<td>D = .21</td>
</tr>
<tr>
<td>Posttest</td>
<td>M = .84</td>
<td>M = .77</td>
<td>M = .91</td>
<td>M = .90</td>
<td>M = .69</td>
<td>M = .85</td>
</tr>
<tr>
<td></td>
<td>D = .26</td>
<td>D = .29</td>
<td>D = .17</td>
<td>D = .18</td>
<td>D = .31</td>
<td>D = .27</td>
</tr>
<tr>
<td>Retention Test</td>
<td>M = .65</td>
<td>M = .56</td>
<td>M = .73</td>
<td>M = .81</td>
<td>M = .60</td>
<td>M = .77</td>
</tr>
<tr>
<td></td>
<td>D = .39</td>
<td>D = .40</td>
<td>D = .34</td>
<td>D = .27</td>
<td>D = .35</td>
<td>D = .30</td>
</tr>
<tr>
<td><strong>Admissible Probability Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>M = -16.11</td>
<td>M = -16.58</td>
<td>M = -17.75</td>
<td>M = -16.68</td>
<td>M = -20.10</td>
<td>M = -13.15</td>
</tr>
<tr>
<td>Posttest</td>
<td>M = 17.34</td>
<td>M = 10.43</td>
<td>M = 23.72</td>
<td>M = 21.28</td>
<td>M = 6.34</td>
<td>M = 18.39</td>
</tr>
<tr>
<td></td>
<td>D = 16.61</td>
<td>D = 17.37</td>
<td>D = 9.49</td>
<td>D = 13.22</td>
<td>D = 17.41</td>
<td>D = 17.81</td>
</tr>
<tr>
<td>Retention Test</td>
<td>M = 5.70</td>
<td>M = 1.58</td>
<td>M = 12.58</td>
<td>M = 17.70</td>
<td>M = 1.18</td>
<td>M = 13.28</td>
</tr>
<tr>
<td></td>
<td>D = 22.71</td>
<td>D = 24.03</td>
<td>D = 19.34</td>
<td>D = 13.31</td>
<td>D = 20.39</td>
<td>D = 19.59</td>
</tr>
<tr>
<td><strong>Time (sec)</strong></td>
<td>M = 676.21</td>
<td>M = 740.00</td>
<td>M = 675.97</td>
<td>M = 725.39</td>
<td>M = 786.22</td>
<td>M = 758.71</td>
</tr>
<tr>
<td></td>
<td>D = 182.52</td>
<td>D = 275.45</td>
<td>D = 301.59</td>
<td>D = 187.26</td>
<td>D = 249.37</td>
<td>D = 187.61</td>
</tr>
<tr>
<td><strong>Attitude Inventory</strong></td>
<td>M = 91.29</td>
<td>M = 92.52</td>
<td>M = 92.12</td>
<td>M = 92.00</td>
<td>M = 93.46</td>
<td>M = 90.64</td>
</tr>
<tr>
<td></td>
<td>D = 15.31</td>
<td>D = 9.05</td>
<td>D = 14.47</td>
<td>D = 7.14</td>
<td>D = 9.77</td>
<td>D = 8.09</td>
</tr>
</tbody>
</table>

Students’ Raw Scores

The mixed factorial design (the split-plot analysis of variance) was used to analyze data from mixed designs – one within-subjects factor (“test”) and one between-subjects factor (“group”). There were three levels of factor in the
within-subjects design in which subjects took all three tests (“pretest”, “posttest”, and “retention test”) and there were 6 levels of factor in the between-subjects design corresponding to the six groups of participants.

The repeated factor, “test,” was a statistically significant main effect for students’ raw score, Wilks’ Lambda = .174, $F_{(2,165)} = 390.784$, $p = .000$, effect size = .826. The interaction between factors “test” and “group” was also a significant main effect for students’ raw scores, Wilks’ Lambda = .848, $F_{(10,330)} = 2.831$, $p = .002$, effect size = .079.

For a follow-up analysis on a significant interaction between the factors “test” and “group”, the ANOVA was used to test for the “group” factor on the levels of the “test” factor.

The ANOVA performed on the pretest data indicated that there was no significant difference in students’ raw scores between the groups, $F_{(5,166)} = 0.570$, $p = .723$, effect size = .017. However, the ANOVA performed on the posttest data revealed that there was significant difference between the groups, $F_{(5,166)} = 4.220$, $p = .001$, effect size = .113. A follow-up post-hoc test, Tukey’s Honestly Significant Difference (HSD) test was conducted to evaluate pair-wise differences among the means. The results of the Tukey’s HSD test indicated that there was a significant difference between the groups (Group B & C and C & E). These results demonstrate that Group C students who received “visual text and added graphics” ($M = 27.44$, $SD = 3.90$) earned significantly higher raw scores than Group B students who received “visual text and added spoken text” ($M = 21.95$, $SD = 7.31$) and Group E students who received “reduced visual text and added spoken text”, ($M = 21.08$, $SD = 6.85$) on the posttest.

From the retention test, the ANOVA data also revealed that there was a statistically significant difference between the groups, $F_{(5,166)} = 3.487$, $p = .005$, effect size = .095. Tukey’s HSD post-hoc test indicated that there was a significant difference between the groups (Group B & D and D & E). These results indicate that Group D students who received “visual text, added spoken text, and added graphics” ($M = 25.42$, $SD = 4.85$) earned significantly higher raw scores than Group B students who received “visual text and added spoken text” ($M = 17.82$, $SD = 9.53$) and Group E students who received “reduced visual text and added spoken text” ($M = 18.42$, $SD = 8.77$) on the retention test.

These results indicate that the students in Group C ("visual text and added graphics") and Group D ("visual text, added graphics, and added spoken text") learned and retained English vocabulary more effectively than students who received the other types of instruction as shown in Figure 3.

![Figure 3. Group’s means for students’ raw scores](image)

**Students’ Mean Degree of Certainty Estimates**

The mixed factorial design was used to analyze the mean degree of certainty estimates. The repeated factor, “test,” was a statistically significant main effect for students’ mean degree of certainty estimates, Wilks’ Lambda = .236,
\( F_{(2,165)} = 267.809, p = .000, \text{effect size } = .764 \). However, the interaction between the factors “test” and “group” was not a significant main effect for students’ mean degree of certainty estimates, Wilks’ Lambda = .908, \( F_{(10,330)} = 1.630, p = .098, \text{effect size } = .047 \).

The ANOVA performed on the pretest (\( F_{(5,166)} = 1.286, p = .272, \text{effect size } = .037 \)) and the retention test (\( F_{(5,166)} = 1.953, p = .088, \text{effect size } = .056 \)) data indicated that there were not statistically significant differences in mean degree of certainty estimates between the groups. However, from the posttest, the results of the ANOVA were statistically significant, \( F_{(5,166)} = 2.910, p = .015, \text{effect size } = .081 \). The results of Tukey’s HSD test also indicated that there was a significant difference in means between the groups (Group C & E). In other words, Group C students who received “visual text and added graphics” (\( M = .91, SD = .17 \)) earned significantly higher mean degree of certainty scores than Group E students who received “reduced visual text and added spoken text” (\( M = .69, SD = .31 \)).

Therefore, the results indicate that students in all groups earned higher scores in general, indicating that their degree of belief probabilities increased when they received multimedia instruction as shown in Figure 4.

**Figure 4.** Group’s means for students’ mean degree of certainty estimates

**Shuford Admissible Probability Scores**

The mixed factorial design was also used to analyze the students’ admissible probability scores. The repeated factor, “test,” was a statistically significant main effect for students’ admissible probability scores, Wilks’ Lambda = .186, \( F_{(2,165)} = 360.841, p = .000, \text{effect size } = .814 \). The interaction between the factors “test” and “group” was also a statistically significant main effect, Wilks’ Lambda = .845, \( F_{(10,330)} = 2.94, p = .002, \text{effect size } = .081 \).

The ANOVA performed on the pretest data indicated that there was no significant difference in admissible probability scores between the groups, \( F_{(5,166)} = 1.222, p = .301, \text{effect size } = .036 \). However, from the posttest, the ANOVA was statistically significant, \( F_{(5,166)} = 4.789, p = .000, \text{effect size } = .126 \). Tukey’s HSD test indicated that there was a significant difference in the means between the groups (Group B & C, C & E, and D & E). These results indicate that Group C students who received “visual text and added graphics” (\( M = 23.72, SD = 9.49 \)) earned significantly higher scores than Group B students who received “visual text and added spoken text” (\( M = 10.43, SD = 17.37 \)) and Group E students who received “reduced visual text and added spoken text” (\( M = 6.34, SD = 17.41 \)). It also shows that Group D students who received “visual text, added graphics, and added spoken text” (\( M = 21.28, SD = 13.22 \)) earned significantly higher scores than Group E students who received “reduced visual text and added spoken text” (\( M = 6.34, SD = 17.41 \)).
For the retention test, the ANOVA was also statistically significant, $F_{(5,166)} = 3.378$, $p = .006$, effect size = .092. Tukey’s HSD test indicated that there was a significant difference in means between groups (Group B & D) on the retention test. The result indicates that Group D students who received “visual text, added graphics, and added spoken text” ($M = 17.70$, $SD = 13.31$) earned significantly higher scores than Group B students who received “visual text and added spoken test” ($M = -1.58$, $SD = 24.03$).

These results lead to the conclusion that there was a significant difference in students’ admissible probability scores between the groups when students received multimedia instruction as shown in Figure 5.

Figure 5. Group’s means for students’ Shuford Admissible Probability Scores

**Time Required to Complete Instruction & Student Attitude Inventory**

The ANOVA showed that there was no significant difference among students who were taught under the six methods of multimedia instruction conditions either in the time required to complete the instruction ($F_{(5, 161)} = 1.070$, $p = .379$, effect size = .032) or in the student attitude inventory ($F_{(5, 163)} = .175$, $p = .972$, effect size = .005).

**Discussion**

*Hypothesis 1: There are no differences in original learning among students who are taught under the six methods of multimedia instruction conditions as measured by raw score, mean degree of certainty estimate, and the Shuford Admissible Probability Scoring Procedure.*

The mixed factorial design (the split-plot analysis of variance) on data obtained from the scores of students who were taught under the six methods of multimedia instruction conditions as measured by the student’s raw scores and the Shuford Admissible Probability Scores on the posttest and retention test showed differences between the six methods of multimedia instruction at the .05 level of significance.

No significant differences were found among treatment groups on data of the students’ mean degree of certainty estimates on the pretest and retention test. Thus, this result leads one to conclude that there were no significant differences among the treatment groups with respect to the degree of certainty of knowledge on the pretest and retention test. However, the results show that, in general, students earned a higher score which indicates that they increased their degrees of belief probabilities when they received multimedia instruction.
The students received higher scores in general, which indicates that they learned better when they received “visual text and added graphics” or “visual text, added graphics, and added spoken text” in their instruction than did students who received other types of instruction (“visual text”, “visual text and added spoken text”, “reduced text and added spoken text”, or “reduced text, added graphics, and added spoken text”). In other words, when visual text was presented with graphics, students may be more motivated to success and achievement in vocabulary learning.

Hypothesis 2: There are no differences in time to complete instruction among students who are taught under the six methods of multimedia instruction conditions.

Results of the ANOVA on data obtained from the students who were taught under the six methods of multimedia instruction conditions as measured by the time required to complete instruction showed no differences between the six methods of multimedia instruction at the .05 level of significance.

Although providing multimedia components required individuals to spend more time to complete the instruction, the amount of time spent was not significant. This result leads one to conclude that there were no significant differences among the six methods of instruction with respect to time needed to complete instruction.

Hypothesis 3: There are no differences in students’ attitude toward instruction among students who are taught under the six methods of multimedia instruction conditions as measured by their score on the student attitude inventory.

Results of the ANOVA on data obtained from the students who were taught under the six methods of multimedia instruction conditions as measured by score on the student attitude inventory showed no differences in attitude toward instruction among students who were taught under the six methods of multimedia instruction conditions. This result leads one to conclude that there were no significant differences among the treatment groups with respect to student attitude.

Based on visual text. The instruction based on visual text (Group A), in general, helped students to learn and retain English vocabulary more effectively than the instruction based on visual text and added spoken text (Group B) and the instruction based on reduced visual text and added spoken text (Group E).

From Mayer and Moreno’s research (2002), the results should have indicated that providing words as narration (spoken text) helped students’ performance more than on-screen text (visual text). However, the results show that written words alone may help students to learn and retain English vocabulary. A possible reason for this is that EFL learners often adopt various strategies to memorize vocabulary words such as word lists or paired associations in which new words are presented with their translations. These strategies often require learners to memorize pair associations directly. Vocabulary learning studies in South Korea have often focused on learning based on visual text alone in printed materials. In other words, written words alone may be better than text and added spoken text because Korean students are more accustomed to memorizing lists.

Added spoken text. The instruction based on visual text and added spoken text (Group B) and based on reduced visual text and added graphics (Group E) reduced students’ ability to learn and retain English vocabulary. According to cognitive load theory (Mayer, 2001; Sweller et al., 1998; Tabbers et al., 2004), the results should have indicated that students could reduce memory load with instruction based on reduced text with added spoken text (Group E). However, in fact, this aid did not help students to learn and retain English vocabulary. This result indicates that neither replacing visual text with spoken text nor adding graphics to the visual text easily generalize to all educational settings.

The lower scores for Groups B (visual text and added spoken text) and E (reduced visual text and added spoken text) may indicate problems in phonic learning. Students may have difficulty in knowing exactly how the words are pronounced. Because many EFL learners in Korea are accustomed to memorizing new words without knowing exactly how they are pronounced, spoken text in the instruction created an unnecessary distraction.

Another possible reason for this is that the test was designed to measure only students’ understanding of a word’s meaning and did not measure their knowledge of the word’s pronunciation. The spoken text seemed to be a
distraction to students who are accustomed to learning a foreign language mainly in the current written text format. Thus, this result leads one to conclude that an effective way to improve learning of new English vocabulary is to avoid the addition of spoken text when explaining what the vocabulary means in Korean.

*Added graphics.* The instruction based on visual text and added graphics (Group C), based on visual text, added graphics, and added spoken text (Group D), and based on reduced text, added spoken text, and added graphics (Group F) helped students to learn and retain English vocabulary more effectively than the other types of instruction as demonstrated on the pretest and retention test. The results show that in general, students earned a higher score, which indicates that they learned better when they received graphics or graphics and spoken text in their instruction than did students who received other types of instruction (Group A, B, and E). In other words, when visual text is presented with graphics, students may be motivated to success and achievement in vocabulary learning.

The likely reason is that text does not usually translate in a manner that is meaningful to the student, while a graphic allows the student to visualize the definition in a more meaningful way. Some words cannot be translated directly and retain meaning for middle school students. When students received the instruction based on visual text only, for example, showing that “taut” means “pulled or stretched tight” may not allow learners to explain what that definition means in ways that make sense to them. The results supported the conclusion that students performed better on vocabulary tests when they were allowed to use a combination of visual text and graphics (Kost et al., 1999).

The results support the concept from dual coding theory (Paivio, 1986, chap. 4; Sadoski & Paivio, 2001, chap 3) that students are likely to build connections between verbal (visual text) and nonverbal (graphics) representations. As Mayer (2001) concluded, we assume that processing of visual text takes place initially in the nonverbal channel and then moves to the verbal channel of the brain. The results appear to indicate that providing both visual text and graphics helped students to “select relevant information, organize it into coherent representations, and integrate it with other knowledge” (Mayer & Moreno, 2002, p. 111) as meaningful learning.

The results appear to indicate that replacing visual text with spoken text and adding graphics to the visual text both do not easily generalize to all educational settings (Tabbers et al., 2004). A possible reason for this is that reduced text was not sufficient to explain the definitions in Korean.

In addition, learning new vocabulary within the context of the instruction based on visual text and added graphics may indicate that

“…students have to make informed guesses as to the meaning of a new word in light of available linguistic cues in the context as well as the relevant knowledge in the learner’s mind, including general knowledge of the world, awareness of the situation, and relevant linguistic knowledge.” (Sun & Dong, 2004, p. 132)

**Conclusion**

This study was an investigation of the use of multimedia components such as visual text, spoken text, and graphics in a Web-based self-instruction program to increase students’ English vocabulary learning at Myungin Middle School in Seoul, South Korea. The findings of the study support the idea that the use of visual media supports vocabulary acquisition and helps increase achievement scores. In particular the results lead one to conclude that an effective way to improve the learning of English vocabulary is to offer graphics to illustrate the definition. Students were likely motivated to success and achievement in vocabulary learning when visual text was presented with graphics because text alone did not usually translate in a manner that is meaningful to the learners, while graphics allowed them to visualize the definition in a more meaningful way.

Vocabulary learning is often used with strategies such as word lists or paired associations in which new words are presented with their translations. These strategies with visual text alone may be outdated and irrelevant to students who are accustomed to visual stimuli and have shorter attention spans. The findings of this study indicate that developers of vocabulary learning instruction and curriculum should reconsider their use of multimedia within their presentations. For example, because presenting too many elements in visual or verbal form can lead to reduced ability to learn and retain vocabulary, visual text, spoken text and graphics must be carefully planned and utilized in the instruction. Replacing text and graphics with spoken text can create an unnecessary distraction for EFL learners
who are accustomed to memorizing new words without knowing exactly how they are pronounced. Reduced text or graphics sometimes does not sufficiently explain the definitions of new words, while spoken text helps developers to save space or time to present messages in the instruction. Developers should select relevant graphics to illustrate the meanings of words with appropriately sized images in multimedia instruction. Integrating text and graphics can allow learners to visualize definitions of the words in a way that fosters meaningful learning outcomes. It can help students to have meaningful learning through the cognitive process, or in other words, “selecting relevant words and images, organizing them into coherent verbal and visual representations, and integrating corresponding verbal and visual representations” (Mayer & Moreno, 2002, p. 111). This has ramifications beyond the world of computer-based learning. Printed instructions and in-class lessons should also be designed to improve learner retention through the implementation of graphics alongside current uses of media.

Future research could focus on measuring students’ knowledge of a word’s pronunciation as well as its meaning. In a test that measures students’ knowledge of the word’s pronunciation, the spoken text may not be a distraction to English vocabulary learning. Other future studies could replicate our methods to discover whether the results would be similar for vocabulary learning for languages other than English and with students from countries other than Korea. Although “all multimedia messages are not equally effective” for EFL students (Mayer, 2001, p. 79), it is hoped these findings can be replicated and expanded for the use of vocabulary learning in other languages and cultures.

References


Training in spatial visualization: The effects of training method and gender

Ahmad Rafi
Faculty of Creative Multimedia, Multimedia University, Cyberjaya, Malaysia // Tel:+603-8312 5549 // Fax:+ 603-8312 5554 // ahmadrafi.eshaq@mmu.edu.my

Khairul Anuar Samsudin and Che Soh Said
Faculty of Information and Technology, Sultan Idris University of Education, Malaysia // khairul@ftmk.upsi.edu.my // chesoh@ftmk.upsi.edu.my

ABSTRACT
This paper reports the findings from an experimental study involving thirty three secondary school students (mean age = 15.5 years) in spatial visualization (SV) training through an interactive desktop virtual environment spatial trainer (iDVEST). Stratified random sampling was used to assign students into two experimental groups and one control group. The first experimental group trained in interaction-enabled DVEST (i-DVEST), the second experimental group trained in animation-enhanced DVEST (a-DVEST), and the control group received conventional training. A multi-factorial pretest posttest design procedure was used and data were analyzed using 2-way ANCOVA. Participants trained in i-DVEST made the highest improvement gain in SV, followed by those trained in a-DVEST, and the control group achieved the lowest improvement gain. In general, male participants achieved higher SV improvement gain compared to female participants. Interaction effect between method of training and gender was observed indicating that male students tended to benefit more when trained in i-DVEST, moderately in a-DVEST, and poorly in conventional method. On the other hand, female students seemed to benefit from training irrespective of the method used.

Keywords
Animation, navigation, visualization, virtual environment, VRML

Introduction

Spatial ability is a cognitive ability that is very important to humans for everyday chores and specialized activities. Simple activities such as arranging furniture in home entail having some knowledge about spatial layout and orientation. The ability to drive on the highway safely requires a driver to be able to make accurate and quick judgment of distances and turns when maneuvering and stopping. In highly technical domains, engineers, architects, and computer-aided designers for example require sophisticated spatial skills and knowledge to understand blueprints and make inferences to the constructions of buildings, components, and artifacts. Without a well-developed spatial ability, a person may encounter serious problems affecting one’s academic pursuit or career.

Spatial ability is conceptualized as a psychological cognitive construct built upon several independent sub-abilities or skills. Linn and Petersen (1985) categorized spatial ability into three components namely spatial visualization, mental rotation and spatial perception. Prior to Maccoby’s and Jacklin’s (1974) study, spatial ability was regarded as an innate ability and thus immune from training effect. Findings from experimental studies thereof have established evidence that this ability is trainable when trainings are designed with specific focus on this ability. One of the important discoveries emerging from studies involving psychometric testing of spatial ability was the revelation of gender differences favoring boys. In general, boys were consistently found to perform better than girls on these tests suggesting that boys generally possess greater spatial skills than girls. Male advantages in spatial ability have been established in reviews by Maccoby and Jacklin (1974), Linn and Petersen (1985), and Voyer, Voyer, and Bryden (1995) where the trends of gender differences were found to be stable and consistent. Their meta-analyses show that the gender differences were highly, moderately and marginally robust in mental rotation, spatial perception and spatial visualization respectively. Interestingly, Maccoby and Jacklin (1974) found that some gender differences in spatial ability initially favored girls at young age, but began to favor boys as individuals approached adolescence. This suggests that environmental or experiential factors may influence the development of spatial ability adding credence to the claim that this ability can be nurtured. Apparently, the explanations of gender differences were premised on biological and environmental factors. Sex hormone (Halpern, 2000; Hier & Crowley, 1982), cerebral lateralization (Gur et. al., 2000; Voyer, Voyer, & Bryden, 1995), and x-linked genetic theory (Harris, 1978; Skuse, 2005) are a few treatises based on biological framework.
On the other hand, several propositions such as differential experience and socialisation (Baenninger & Newcombe, 1989; Quaiser-Pohl & Lehmann, 2002; Voyer, Nolan, & Voyer, 2000) and gender-role identification (Signorella & Jamison, 1986; Massa, Mayer, & Bohon, 2005) were fixated on environmental perspective attributing to social-cultural aspects. A new approach of interpretation has been advocated converging on these two perspectives where spatial performance differences were conceived to originate from an interaction of biological and experiential factors. Theories by Sherman (1978) and Casey (1996a, 1996b) supported this interaction by reporting on innate predisposition for spatial abilities that leads to self-selection of activities that ultimately influences the development of a person’s spatial ability or skills. Irrespective of the plethora of theories having substantial claims, many have agreed that experiential factors are detrimental in producing gender differences. Greater participation in spatial activities resulted in higher performance in spatial task, and both genders improved after spatial training (Baenninger & Newcombe, 1989; Olkun, 2003; Rafi, Khairulanuar, & Ismail, 2006). Of late, training in spatial ability has adopted new and novel technologies in particular virtual reality creating an interactive 3D computer-generated environment. Training became more efficient and effective through visualization, animations and greater interaction with virtual training objects (Mantovani, 2001; Moyer, Bolyard, & Spikell, 2001; Rafi, Khairulanuar, Haniff, Maizatul, & Mazlan, 2005). The focus of this study was on spatial visualization that is critical in technical education using the desktop virtual reality.

Theoretical considerations in developing the training application

The topics of spatial training have been deliberated extensively in the literature covering interventional programs or applications developed to train users in spatial tasks. However, some of these training programs were generic and lacked sound theoretical framework for efficacious use. In this study, the authors adopted a theoretical approach informed through a review on the literature of current theories of learning and psychometric testing. The former will guide the development of instructional setting or training environment that helps learners maximize their learning or training potentials whilst the latter will ensure that relevant spatial tasks are developed using appropriate learning objects or training tools than tap on spatial ability. Pairing these two factors together will facilitate the development of an effective spatial training application.

The training application was designed based on the framework espoused by the constructivist view of learning. Dalgarno (2001) outlined three broad principles of this view based on studies by Kant (1946) and Dewey (1938) for the first principle, Piaget (1969) for the second principle, and Vygotsky (1978) for the third principle. The first principle informs that individuals formed their own representation of knowledge and there was no particular ‘correct’ representation of knowledge. The second principle prescribes that learning occurred when learners uncovered a deficiency in their knowledge representation or an inconsistency between their current knowledge representation and their experience during active exploration. Learning that takes place within a social context is the third principle of this constructivist view. These three broad principles have been adopted in formulating approaches to the teaching and learning process. To explicate the intricacies of the interpretation, Dalgarno (2001) cited Moshman’s (1982) categorization of constructivism into three distinct classes namely endogenous, exogenous, and dialectical constructivism. Endogenous constructivism emphasizes the individual nature of each learner’s knowledge construction process relegating the role of teachers to facilitators. Exogenous constructivism highlights the formation and refinement of knowledge representations through learning by instructions with support from exercises entailing active cognition. Dialectical constructivism supports learning through realistic experience coupled with social interactions among teachers, experts and peers providing the essential scaffoldings.

The appropriate use of training or learning objects was informed by reviewing the items in the psychometric test battery that are utilized as part of intelligence and employment assessments. These test batteries have been evaluated showing strong validity and reliability measures for the cognitive-psychological constructs such as spatial visualization, mental rotation and spatial perception. The development of the training objects was based on test items under the category of ‘count touching blocks’ that measures the ability to visualize forms in space and to manipulate them mentally (Wiesen, 2004). This ability is also considered a strong component of mechanical aptitude deemed vital in learning technical and industrial fields.
The development of an interactive Desktop Virtual Environment Spatial Trainer (iDVEST)

The training program comprises training objects as part of the instructional learning activities in a platform designed and adapted from the principles or views of constructivist learning theory. A desktop virtual environment was chosen capitalizing on two important criteria; three-dimensionality and practicality. The former is essential in engendering 3D-space learning environment that facilitates better understanding on spatial properties and relationships of objects and space. The latter is more concerned with the ease of development, implementation and maintenance in typical Malaysian schools. The terms virtual reality (VR) and virtual environment (VE) are interchangeably used to describe a computer-generated three-dimensional (3D) environment allowing real-time interactions by means of one of more control devices and involving one or more sensorial perceptions (Ausburn & Ausburn, 2004; Schneiderman, 1993). Though lacking the immersion factor associated with high-end immersive virtual reality systems, desktop virtual environment has made significant impact in learning and training given the continually increasing processing capability of desktop computers providing faster and better graphics delivery. Desktop VR technology emerged in formats such as a mouse-controlled navigation in a 3D-environment that is affordable especially for general classroom usage (Ausburn & Ausburn, 2004). Realism, flexibility, interactivity, and easy learner control of the screen-based environments offer great learning potentials for users in the desktop VR (Shneiderman, 1993). The development of open-standard, non-proprietary Virtual Reality Modeling Language (VRML) (Beier, 2004) has added wider adoption of desktop VR that is web-deliverable over the Internet for greater distribution.

The development of an interactive desktop virtual environment spatial visualization trainer (iDVEST) was carried out as a web-based application that constitutes a series of exercises pertaining to spatial visualization tasks involving the configuration of stack of blocks. Each exercise is presented with a question requiring a learner to count the virtual blocks that are in contact with a virtual target block. Interactions with the virtual objects are performed through a viewer application viz. Parallel graphic’s Cortona VRML client. Animations of the virtual objects were also programmed along specified paths in iDVEST as a cognitive tool for assisting learners in visualization and provision of correct spatial solutions. The interaction mode of the former and animation mode of the latter were operationally categorized as the two experimental treatments namely i-DVEST and a-DVEST training conditions respectively. Feedback mechanism was also added to acknowledge learners the incorrect or correct responses facilitating the correction of mistakes and development of new plans.

![Figure 1: The iDVEST comprising VRML client interface, virtual objects and questions](image)
Dalgarno (2001) advised judicious use of elements from each view to produce an efficacious training environment. Exercises are in the form of a series of questions requiring learners to read carefully the instructions hinged on exogenous constructivist view emphasizing the role of direct instruction. Interacting with the virtual objects to observe spatial configuration from multiple views helps learners construct and refine their mental models reflected the endogenous constructivist view. The combination of the feedbacks to inform learners the correct or incorrect status of the responses and animations constitutes a scaffolding mechanism in line with the dialectical view of constructivism helping them to arrive at the correct solutions. Figure 1 shows the layout of the iDVEST containing the questions and virtual training objects.

**Purpose of study**

Spatial ability or skills such as spatial visualization are essential cognitive abilities in secondary education especially for students enrolled in technology education. Investigation on spatial visualization improvement based on method of training and gender was carried out in this study with emphasis on active exploration and interaction in the interactive desktop virtual environment. The study sought to compare the effectiveness of spatial training using three methods of training conditions namely interaction-enabled desktop virtual environment (i-DVEST), animation-enhanced desktop virtual environment (a-DVEST), and conventional method. Three research questions to address the issues in the study are as follows:

a) Is there a significant improvement in participants’ spatial visualization after training?

b) Are there significant differences in improvement gain of spatial visualization based on the method of spatial training?

c) Is gender a significant factor in influencing the training outcome after the participants trained in spatial training?

Three hypotheses were formulated to answer the three research questions pertaining to spatial visualization improvement, effectiveness of training method, and influence of gender. The three hypotheses are as follows:

a) Participants’ spatial visualization will improve significantly after spatial training.

b) Improvement gain in spatial visualization will be higher in i-DVEST training, moderate a-DVEST training, and least in conventional training.

c) Spatial Visualization improvement gain will be greater for male students compared to their female counterparts.

**Methods**

**Participants**

A class of technical education program comprising thirty-three secondary school pupils (13 girls and 20 boys, mean age = 15.5 years) of SMK, Kuala Kubu, Selangor, Malaysia volunteered to participate in this experimental study. Stratified sampling rather than simple random sampling was used as the sample did not have equal number of genders where the student class comprised approximately 40% girls and 60% boys. Proportionate allocation technique ensured that randomized assignment of students into groups would reflect the same male and female strata. Three groups were formed namely two experimental groups and one control group. The first experimental group (4 girls and 7 boys) trained in the interaction-enabled desktop virtual environment (i-DVEST), the second experimental group (4 girls and 7 boys) received training in animation-enhanced desktop virtual environment (a-DVEST), and the control group (5 girls and 6 boys) was exposed to conventional training.

**Instrument and Instructional Materials**

The spatial trainer developed by the authors contains spatial tasks specifically focusing on mental exercises tapping on spatial visualization. It is an integrated web-based application developed on desktop virtual environment containing virtual objects that serve as training objects for the spatial exercises. Instructional information were organized with questions and presented on the lower half of the screen and objects for training were presented on the top half of the same web page. The spatial trainer contains features that support training namely interaction via mouse-controlled VRML client, animations and response feedbacks. The first training condition was set to allow only navigation and exploration (operationalized as interaction) with the virtual objects and response feedbacks. The
second experimental condition involved training with animation and response feedback features. The control condition employed similar questions of the spatial exercise in the conventional way (i.e. printed material) supplemented with static graphics.

A computerized spatial visualization test based on the Spatial Visualization Test (Middle Grade Mathematical Project, 1983) was also developed by the authors to measure participants’ spatial visualization prior to and after training providing the pretest and posttest measures of this ability. The use of this instrument was qualified by the Cronbach’s reliability coefficients ranging from .72 to .88. The test comprises 32 multiple-choice items, which is to be completed in 15 minutes. Each item involves an object made up of small unit cubes seen from a certain perspective. Participants are required to determine which one of the five other objects is the same as the one depicted, but from another view. Scoring is based on the number of correct answers out of 32 possible and converted this to a percentage. In addition, a 12-item questionnaire with one open-ended question was administered to every student after the completion of the training. This instrument measured students’ perceptions on essential training factors namely training effectiveness, motivation and system usability (applicable to i-DVEST only). Each item response was rated along a Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly disagree) yielding a maximum 20 points for each factor. The open-ended question helped reveal greater insight about the training experienced by the students. A professor who has expertise in educational research reviewed the questionnaire and the researchers used reflective analysis to evaluate a phenomenon relying on intuition and judgment as recommended by Gall, Borg and Gall (2003). All the comments were translated from Malaysian language to English by the researchers.

Procedure

A multi-factorial pre-test post-test experimental (3 x 2) design procedure comprising two independent variables namely training method (3 levels: i-DVEST, a-DVEST, and control) and, gender (2 levels: girls and boys). The dependent variables in the study were the spatial visualization mean scores providing both the pretest and posttest measures. The training program was conducted for four weeks with the first used for computerized pretest on spatial visualization, administration of consent forms, briefing and familiarization of the training application. Actual training began on the second week involving two labs for the experimental groups and a classroom for the control group. Table 1 shows the training activities spanning the 4-week duration.

<table>
<thead>
<tr>
<th>Week</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pretesting, administration of consent form and task familiarization.</td>
</tr>
<tr>
<td>2</td>
<td>Spatial visualization training of low tasks.</td>
</tr>
<tr>
<td>3</td>
<td>Spatial visualization training of moderate task.</td>
</tr>
<tr>
<td>4</td>
<td>Spatial visualization training of complex task, posttesting, and administration of questionnaire</td>
</tr>
</tbody>
</table>

Each training session lasted for two hours with first session dealing with easy spatial exercises, moderately difficult exercises on the second session, and complex spatial tasks on the third session. Each exercise requires the participants to determine the correct solution among six choices for a given task (i.e., the correct number of blocks touching a target block contained in a stack of blocks). The authors and their subject teacher assisted in the supervision throughout the training sessions.

Participants in the i-DVEST condition initially worked through the spatial exercises by following the instruction posted in the questions and then interacting with the virtual objects via mouse control on the VRML client interface. The interactions involved panning, twisting, rotating, and rolling of the virtual stack of blocks providing multi-point viewing that enhances visual perception. Participants in the a-DVEST group performed the same spatial exercises as the first group except that the interaction mode was disabled. However, they could use the animation button to launch the animations to help them solve the spatial exercises. Each animation highlights individual blocks in contact with a target block and then shows the movement of the blocks one by one away from the target block (see Figure 2).
This effectively suspends the blocks in space effectively exposing the target block in its canonical position facilitating visualization of the tasks. The control group followed through the training by performing similar spatial exercises presented in printed materials containing similar instructional activities. Solutions to the spatial exercises comprise static graphics that show exploded views of the stack of blocks helping them in the visualization of the spatial configuration. The posttest was administered immediately after the completion of the final training session using the same computerized spatial visualization test. A short survey questionnaire was then administered to elicit information regarding participants’ perception on the effectiveness of the training they received.

**Findings**

The participants’ levels of spatial visualization (SV) were measured using the appropriate spatial measurement instruments prior to spatial training. The scoring procedure for SV was based on the percentages of correct responses. The mean score of the pretest for girls \( (n = 13) \) and boys \( (n = 20) \) were 42.38 and 47.55 summarized in Table 2.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Spatial Visualization</td>
<td>13 42.38</td>
<td>15.26</td>
</tr>
</tbody>
</table>

Differences attributed to gender factor in spatial visualization were investigated prior to the treatments of spatial training. An independent samples t-tests were performed to detect if there were any gender differences among the participants in the study. For the spatial visualization pretest, the difference in mean scores of male and female participants was not statistically significant, \( t(31) = -.94, p = .36 \), establishing that the participants’ spatial visualization were equivalent prior to training. Descriptive statistics and univariate analysis of variance (ANOVA) were conducted to detect if there were significant differences among the three groups. The ANOVA procedure performed used the means scores of SV as the dependent variables and condition of training as the independent variables. Table 3 summarizes the descriptive statistics of the three groups.

<table>
<thead>
<tr>
<th>Measures</th>
<th>i-DVEST</th>
<th>a-DVEST</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n Mean  SD</td>
<td>n Mean  SD</td>
<td>n Mean  SD</td>
</tr>
<tr>
<td>Spatial visualization</td>
<td>11 53.36 17.19</td>
<td>11 40.00 12.69</td>
<td>11 45.52 15.46</td>
</tr>
</tbody>
</table>
For the pretest measure of SV, the 11 participants in the i-DVEST group had a mean of 53.36 (SD = 17.19); the 11 participants in the a-DVEST group had a mean of 40.00 (SD = 12.69), and the 11 participants in the control group had a mean of 45.52 (SD = 15.46). Table 4 summarizes the result of the analysis of variance for SV measure in the three groups.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1072.061</td>
<td>2</td>
<td>536.030</td>
<td>2.445</td>
<td>.104</td>
</tr>
<tr>
<td>Within Groups</td>
<td>6576.182</td>
<td>30</td>
<td>219.206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7648.242</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There were no significant differences between these groups prior to spatial training, \( F(2,30) = 2.45, p=.10 \). In other words, all the groups on the onset of this study were statistically equivalent in spatial visualization. After spatial training, the 33 participants had an average difference from pretest to posttest spatial visualization scores of 6.56 (SD = 6.54), indicating that the spatial training had improved their spatial visualization levels, \( t(32) = 5.76, p = .001 \) thus supporting the first hypothesis of the study.

Two-way analysis of covariance (2-way ANCOVA) was utilized to test the main effects of training condition and gender on spatial visualization, controlling the effects of prior spatial visualization. This statistical procedure also tested the interaction between training condition (i-DVEST, a-DVEST, control) and gender (girl, boy). The dependent variables, covariates, and independent variables were SV posttest measurements, SV pretest measurements, and gender and training condition respectively. The appropriateness of using this analysis was tested by first conducting the analysis using statistical model containing interaction terms between the covariates (i.e., the pretest mean scores of spatial visualization) and the independent variables to assess the assumption of homogeneity of slopes. The analysis of the ANCOVA to test this assumption is summarized in Table 5.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training condition</td>
<td>84.176</td>
<td>2</td>
<td>42.088</td>
<td>1.461</td>
<td>.251</td>
</tr>
<tr>
<td>Gender</td>
<td>7.917</td>
<td>1</td>
<td>7.917</td>
<td>.275</td>
<td>.605</td>
</tr>
<tr>
<td>Pre_Spatial Vis.</td>
<td>4047.715</td>
<td>1</td>
<td>4047.715</td>
<td>140.503</td>
<td>.000</td>
</tr>
<tr>
<td>Pre_Spatial Vis*Condition.</td>
<td>8.883</td>
<td>2</td>
<td>4.441</td>
<td>.154</td>
<td>.858</td>
</tr>
<tr>
<td>Pre_Spatial Vis*Gender.</td>
<td>29.735</td>
<td>1</td>
<td>29.735</td>
<td>1.032</td>
<td>.319</td>
</tr>
<tr>
<td>Error</td>
<td>720.218</td>
<td>25</td>
<td>28.809</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8241.061</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The interaction of the covariates (i.e., prior spatial visualization) with training condition that was used to test for homogeneity of regression for the spatial visualization test was not statistically significant, \( F(2,25) = .15, p = .86 \). The homogeneity of regression test for the interaction between the covariate and gender was also not statistically significant, \( F(1,25) = 1.03, p = .32 \). This indicated that the assumption of parallelism of slopes was met supporting the use of ANCOVA.

For the posttest measure of SV, the 11 participants in the i-DVEST group had an adjusted mean of 55.74 (SD = 16.83); the 11 participants in the a-DVEST group had an adjusted mean of 51.59 (SD = 11.44), and the 11 participants in the control group had an adjusted mean of 47.26 (SD = 12.83) as summarized in Table 6.

Results of the ANCOVA revealed a statistically significant main effect for training condition, \( F(2,26) = 15.53, p = .001 \) as shown in Table 7. Participants trained in i-DVEST outperformed others and participants in the a-DVEST were better than control group. To determine where the differences among the training methods were, Bonferroni’s Post Hoc Test was employed to test for significance. All tests were conducted using the adjusted means, controlling for any differences in prior spatial visualization ability.
Table 6: Observed Means, Adjusted Means, and Standard Deviations for Spatial Visualization after training by gender and training condition

<table>
<thead>
<tr>
<th>Gender</th>
<th>i-DVEST</th>
<th>a-DVEST</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Females</td>
<td>4</td>
<td>52.75</td>
<td>15.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(48.96)</td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>7</td>
<td>70.79</td>
<td>14.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(62.53)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>obs. means</td>
<td>11</td>
<td>64.23</td>
</tr>
<tr>
<td></td>
<td>adj. means</td>
<td>(55.74)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Adjusted means are presented in parentheses

Table 7: Analysis of Covariance of Spatial Visualization after training

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre_Spatial Vis.</td>
<td>4530.639</td>
<td>1</td>
<td>4530.639</td>
<td>394.957</td>
<td>.000</td>
</tr>
<tr>
<td>Condition.</td>
<td>356.262</td>
<td>2</td>
<td>178.131</td>
<td>15.528</td>
<td>.000**</td>
</tr>
<tr>
<td>Gender.</td>
<td>70.836</td>
<td>1</td>
<td>70.836</td>
<td>6.175</td>
<td>.020*</td>
</tr>
<tr>
<td>Condition * Gender.</td>
<td>453.209</td>
<td>2</td>
<td>226.605</td>
<td>19.754</td>
<td>.000**</td>
</tr>
<tr>
<td>Error.</td>
<td>298.252</td>
<td>26</td>
<td>11.471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total.</td>
<td>8241.061</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

The post hoc test for the training condition variable was tested at the pre-established alpha level of .05. The i-DVEST condition was compared to the a-DVEST condition revealing a mean difference of 4.156 and a significance of .048 that indicated that there was significant difference between the two training methods as depicted in Table 8.

Table 8: Bonferroni Post Hoc Test results by training condition

<table>
<thead>
<tr>
<th>(I) Training condition</th>
<th>(J) Training condition</th>
<th>Mean Difference (I-J)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental 1</td>
<td>Experimental 2</td>
<td>4.156*</td>
<td>.048</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td>8.484**</td>
<td>.000</td>
</tr>
<tr>
<td>Experimental 2</td>
<td>Experimental 1</td>
<td>-4.156*</td>
<td>.048</td>
</tr>
<tr>
<td>control</td>
<td></td>
<td>4.328*</td>
<td>.022</td>
</tr>
<tr>
<td>Control</td>
<td>Experimental 1</td>
<td>-8.484**</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Experimental 2</td>
<td>-4.328*</td>
<td>.022</td>
</tr>
</tbody>
</table>

*p<.05, **p<.001.

The comparison of the i-DVEST condition with the control condition revealed a mean difference of 8.484 indicating a significant difference (.001) between the two types of training. A mean difference of 4.328 with significance level of .022 was found when comparing the a-DVEST and control training methods indicating a significant difference between participants’ spatial visualization in the two groups. Clearly, the post hoc test showed there were spatial visualization differences on all the three training conditions.

Similarly, there was significant main effect of gender, $F (1,26) = 6.18, p = .02$, favoring boys as shown in Table 8. The interaction between training condition and gender was found to be statistically significant, $F (2,26) = 19.75, p = .001$, indicating that the effectiveness of training method varied depending on participants’ gender. Figure 3 illustrates the trend analysis of adjusted mean spatial visualization performance scores by training condition and gender.

134
The intersection of the slopes shows that boys were better at spatial visualization tasks than girls for the i-DVEST condition. Both genders were equivalent in their spatial visualization after training in the a-DVEST condition. The same high performance of boys was not replicated in the control condition. In fact, female participants were better than their male counterparts were when trained using the traditional method as shown in Figure 3. In addition, data elicited from the questionnaire provides additional quantitative data (from the 12-item questions) and qualitative data (from the open-ended question). The followings are the mean scores for perceived training effectiveness indicating improved visualization: 13.50, 16.43, and 15.36 (for i-DVEST); 14.50, 16.71, and 15.91 (for a-DVEST); 14.40, 14.67, and 14.55 (for conventional). The followings are the mean scores for perceived motivation after training as reported by girls, boys, and all students: 12.25, 16.57, and 15.00 (for i-DVEST); 15.50, 16.29, and 16.00 (for a-DVEST); 13.60, 12.00, and 12.73 (for conventional). The mean scores for system usability of i-DVEST as perceived by the girls, boys and all respondents are 12.75, 15.57, and 14.55 respectively.

Ten of the thirty-three participants answered the open-ended question involving 5 participants (3 girls, 2 boys) in i-DVEST, 3 participants (2 girls, 1 boy) in a-DVEST, and two girls in the control group. Most of them indicated that their visualization has improved after going through the spatial visualization exercises.

**Discussion**

**Potential gender and group differences prior to spatial training**

Prior to training, all groups were equivalent in spatial visualization ability ensuring that resultant differential outcomes of the spatial training were attributed to the effects of the training methods. Gender parity was also observed where both boys and girls were equivalent in spatial visualization. This was not unexpected given that gender differences are minimally small in spatial visualization compared to robust differences in other spatial abilities especially in mental rotation (Halpern, 2000; Linn & Petersen, 1995; Maccoby & Jacklin, 1974). Another plausible reason is that boys and girls were from similar academic backgrounds that correlate well with their spatial abilities. Students involved in this study were drawn from a class of technology education program having good mathematics grades as one of the prerequisites for technical studies prescribed by the Malaysia’s Ministry of Education policy (KPM, 1998). Significant relations between mathematics achievements and spatial ability measures have been established suggesting that high spatial individuals were better in mathematics compared to low spatial individuals (Baenninger & Newcombe, 1995; Kulp, Earley, Mitchell, Timmerman, Frasco, & Geige, 2004; Tartre,
It is highly probable that there were no superiority of spatial abilities of one gender over the other since both genders were of similar level in mathematics. Seeing from a wider perspective, this finding may not be unique as gender differences on cognitive and psychosocial tasks are getting smaller and declining (Hyde, 2005; Linn & Hyde, 1989).

The effects of training method and gender on spatial visualization training

Comparison of mean scores of spatial visualization before and after spatial training for all groups indicates that the participants have considerably improved their spatial ability. This suggests that spatial training in the three training have improved their spatial visualization thus lending support for the first hypothesis. This finding was further qualified from the data elicited from the self-reported questionnaire indicating all participants did agree that their visualization have improved after engaging in spatial exercises. Additionally, the qualitative data helped explain this result. The major responses of the participants particularly i-DVEST for the open-ended question revealed that interacting with the training objects enables better comprehension of the spatial configuration of stacked blocks in space. Phrases such as “I can see more clearly from this view”, “It helps when I picture it in my mind”, “Viewing from several angles is helping me”, and “rolling and twisting the objects exposes hidden faces” are indicative of better visual perception leading to improved visualization.

Further examination revealed that there was a significant main effect of treatment condition of spatial training thus supporting the second hypothesis. Post hoc test analysis revealed that the greatest performance difference was between i-DVEST group and control group. The difference in spatial visualization between i-DVEST group and a-DVEST group was also quite substantial. The a-DVEST group was better than the control group indicating that spatial training was less effective in the conventional training method. The statistical analysis also lends support for the third hypothesis favoring the boys. In addition, there was a substantial interaction between method of spatial training and gender was observed where a pattern emerged indicating the convergence and divergence of performance measurements of girls and boys respectively (see Figure 4). Boys tended to gain differential level of improvement gain in spatial visualization based on different training conditions whilst girls were more likely to improve irrespective of training method used.

Spatial training in i-DVEST condition

The effectiveness of i-DVEST compared to the other two methods of spatial training only manifested for male participants. Girls in the same group were comparable insofar to the other two groups. In fact, when comparing the means of spatial visualization posttest scores, girls in the experimental group attained the lowest performance across all groups. The high degree of interaction in the i-DVEST had benefited male participants since this feature was absent in other types of training. In solving the spatial visualization tasks, users were able to interact with the virtual objects by panning, twisting and rolling in the interactive desktop virtual environment giving multiple viewpoints thus facilitating the visual perception and processing of objects in 3D space. Learners could view these objects from close up or from a distance when examining specific and holistic features of the artifacts respectively concurring with Smith’s (2001) study that suggested alternating between interaction and observation was the best way to learn spatial visualization. Although there were no significant gender differences in spatial visualization, boys were more superior then girls in this cognitive ability influencing the effort put in when learning in i-DVEST. Heeter (1994) found that boys generally were more interested in VR learning experiences with interaction compared to girls. Similarly, Ziemek’s (2006) indicated that gender has an influence on how attracted an individual is to the electronic games. Three-dimensional (3D) and two-dimensional (2D) electronic games are most appealing to boys and girls respectively. Boys probably were more adept at moving around in virtual environment and interacting with the virtual objects due to their superior spatial ability compared to girls. Observations during the training paralleled her finding indicating that boys were far more active in exploring i-DVEST than girls were. This gender influence was qualified by the qualitative data from two female respondents’ comments (e.g., “need time to get used to using the interface” and “it’s not that straight forward to have a good scene”) pointing out that navigating the virtual training environment can be quite daunting for them. In contrast, a male respondent’s comment (e.g., “I found it engaging to do the exercises in this setting”) seemed to reinforce the gender gap in virtual environment. From the constructivist perspective, the training process seemed to have helped boys trained in i-DVEST to continually construct and refine the correct mental schema of the solutions of the spatial problems. Most revealing comment (e.g., “If I committed an error or got confused, I can get back to and navigate the scene till it becomes more meaningful”) was from one male...
respondent indicating that when he committed an error, he could always return to view the training objects from other directions thus avoiding further errors. Clearly, this learner’s comment demonstrates evidence of accessing prior experience and comparing his flawed understanding to new experience in the training environment indicative of a constructivist training. This mode of training effectively allows him to revise and refine his imperfect mental model by making correct inferences resulting in enriched experience. Additionally, participant motivation elicited from the questionnaire indicated that boys rated the novel training higher than girls did. Studies have shown that motivation influenced the level of engagement in task thus affecting performance.

Spatial training in a-DVEST condition

The second most effective in spatial visualization training was the condition of training in a-DVEST with animation features. Participants utilizing these features gained better insights and understanding of tasks that demand visual perception and cognitive processing based on spatial visualization. Animations were used to fulfill a cognitive function where in this role they support students’ cognitive processes that ultimately make them understand the subject matter better (Love, 2004). Qualitative data from two girls’ responses (e.g., “animations help!” and “the steps are clear”) suggest that animations serve as a cognitive tool to facilitate them seeing the proper steps that they could replicate in their minds during problem solving. These two girls and a boy indicated that they found the animated feature to be helpful especially in solving difficult tasks. The set of spatial tasks designed in this study requires the participants to solve problems associated with stack configurations composed of regular rectangular block. In determining the number of blocks making contact with a target block, the participants have to visualize the blocks in terms of their positions and relations in the 3D space. Animations for the relevant virtual blocks are launched when each will be highlighted (i.e., changing into flashing color) in its position in relation to the target block and then moves away from the target block. This allows participants to observe the target block positioned in space and its relation to the moved blocks. Apparently, the effect of these animations might have facilitated the externalization of the internal visualization process that the participants would invoke in solving the spatial visualization tasks that consequently lessens the cognitive effort in problem solving. Spatial training enhanced with animations seems to benefit both male and female participants in this study where their performances in this group were equivalent.

Spatial training in conventional condition

Another interesting trend in the interaction between training condition and gender was observed in the control group. Girls in this group exhibited higher spatial visualization compared to boys. In fact, male participants in this particular group showed the lowest improvement in spatial visualization among all groups that warrants some explanation to this unexpected finding. The conventional training required participants to work through a similar set of questions in printed exercise worksheets complemented with solutions of the tasks consisted of ‘exploded views’ of the blocks. This type of training was found to advantage boys since earlier study by Geiger and Litwiller (2005) noted that gender differences were found for question answering information from diagrams but not for textual information. The fact that male advantage was not replicated in this setting may be attributed to motivational rather than cognitive construct as indicated by low perceived motivation in the self-reported questionnaire. Girls’ motivational scores of the questionnaire were almost similar with other training conditions reflecting consistent interest and level of effort they invested in training. However, the boys reported lower perceived motivation as they reported learning based on printed materials were less appealing consequently undermining effort to practice. In addition, girls were observed to spend more time practicing on the task compared to boys.

Conclusions

Several important findings emerged from the discussion revealing insights in spatial visualization training are summarized as follows:

a) Spatial visualization training in interaction-enabled desktop virtual environment (i-DVEST) was the most effective where the participants had made substantial improvement. Interactions with the virtual objects provides better visual perception of spatial arrangements of the spatial tasks that leads to better visualization especially for boys. On the other hand, the lack of prior computer experience or familiarization and a slight disadvantage in spatial ability might have confounded girls’ performance in training requiring a high degree of interaction and
navigation. Consequently, girls might have not experienced better visual cues and visualization of the task that led to higher cognitive effort to imagine these spatial configurations leading to poor problem solving ability.

b) Animation-enhanced desktop virtual environment (a-DVEST) training was the second most effective in spatial visualization. Animations of the tasks provide the perceptual cues of the spatial arrangements of the spatial problems leading to improved visualization. Both genders benefited in this training setting indicating that animation helped minimize the cognitive effort to visualize the spatial tasks. Training with animations seemed less cognitively challenging that the above method of training.

c) Conventional training relying on printed materials was least effective among the three methods of training. Poor representation of training objects of 3D into 2D dimensional format minimizes cognitive correspondence between the two representational methods hindering visualization of the spatial tasks. Girls gained relatively higher SV improvement than boys that might be partly attributed to motivational factor as the latter were found to be less enthusiastic in training in this condition. Lack of motivation induced less commitment to engage in training leading to poor performance.

The outcomes of the research may contribute positive educational values in actual teaching of courses in technology education that rely on spatial intelligence or spatial skills. This is particularly importance in the learning of technical graphics (i.e., engineering drawing), computer aided design (CAD), engineering design, and mathematics. Significant positive correlation between technical graphics and spatial visualization has been consistently reported in the literature of spatial training indicating the strong emphasis of developed skills of the latter for successful learning of the former. Students with low spatial ability namely spatial visualization taking technical graphics are most likely at risk in learning the subject matter thus impeding their success. Interventional or remedial programs can be instituted employing technology-driven learning or training tools such as iDVEST to help improve students’ spatial visualization.

Technical graphics course conducted in selected Malaysian schools encompasses one-and-half hour teaching session followed by a 2-hour lab practice session per week. The application of such tools can be embedded into normal lab practice sessions for spatial training of 30-45 minute duration spaced evenly throughout a semester enabling low spatial students to engage in constant practice to help develop their spatial skills. Through this gradual development of spatial skills, these students will be able to learn the subject matter with greater efficacy as they may have attained the proper and efficient strategy in solving technical graphic tasks or problems that are spatial in nature.

The applicability of this study is quite limited given the size of the sample used. Inevitably, the study used the only available class of technology education study in the selected school that is typical in most Malaysian schools. In addition, the authors were only allowed four weeks to conduct the study where gender differences were observed. Future research should use substantial sample size and longer training duration to improve generalizability and address gender difference respectively. Overall, the study has shed some light regarding the effectiveness of an interactive desktop virtual environment as a training platform for spatial visualization. High degree of interaction coupled with programmed dynamic behaviors of virtual objects and feedback mechanism can enrich the training process. However, a cautionary approach to utilizing this training method is advised. An intricate interplay between gender and method of training has been observed entailing careful selection of appropriate objects and tasks, method of instructional delivery and training environment in spatial training.

References


Middle Grade Mathematical Project (1983). *Spatial visualization test*, Department of Mathematics, Michigan State University.


Competence Description for Personal Recommendations: The importance of identifying the complexity of learning and performance situations

Frans J. Prins
Faculty of Social and Behavioural Sciences, Department of Pedagogical and Educational Sciences, Research Centre Learning in Interaction, Utrecht University, The Netherlands // f.j.prins@uu.nl

Rob J. Nadolski, Adriana J. Berlanga, Hendrik Drachsler, Hans G.K. Hummel and Rob Koper
Educational Technology Expertise Centre, Open University of the Netherlands, The Netherlands // rob.nadolski@ou.nl // adriana.berlanga@ou.nl // hendrik.drachsler@ou.nl // hans.hummel@ou.nl // rob.koper@ou.nl

ABSTRACT
For competences development of learners and professionals, target competences and corresponding competence development opportunities have to be identified. Personal Recommender Systems (PRS) provide personal recommendations for learners aimed at finding and selecting learning activities that best match their needs. This article argues that a clear-cut description of the concept of ‘competence’ is needed for appropriate system-based personal recommendations. It is proposed to extend current initiatives on standardization of modelling competences with the concept of ‘Learning and performance situation’ (LP-situation) and it is claimed that this extension has added value for personal recommendations for adequate selection of learning activities, for determining proficiency levels of competences, for the design of learning tasks, and for accreditation of prior knowledge. Advantages and disadvantages are discussed.

Keywords
Competence description, learning and performance situations, personal recommender system

The concept of competence is strongly associated with post-secondary education (e.g., Mulder, Wesselink, Biemans, Nieuwenhuis, & Poell, 2003; Stoof, Martens, & Van Merriënboer, 2007; Westera, 2001) as well as professional development (e.g., Eraut, 1994). Many work organisations and educational institutes use the concept of ‘competence’ for describing performance ability for particular occupations or jobs or for describing educational objectives. For instance, in the Netherlands, the competence requirements of good quality teachers are classified in seven competences: interpersonal competence, pedagogical competence, subject knowledge & methodological competence, organizational competence, competence for collaboration with colleagues, competence for collaboration with the working environment, and competence for reflection and development (SBL, 2004). Similarly, psychology students of the Open University of the Netherlands have to acquire three competences: (1) research competence, (2) diagnosis competence, and (3) intervention competence.

For competences development of learners and professionals, target competences and corresponding competence development opportunities have to be identified. Thereupon, learners may acquire the target competences by participating in so-called Competence Development Programs (CDPs). A CDP is an ordered set of learning activities, either formal or informal, that are used to build competence in a certain discipline or job (Herder et al., 2006; Koper, 2006). An example of a CDP is a sequenced set of courses to be followed in order to get a Master of Science degree in psychology. Currently, many formal and informal CDPs exist, from different educational institutes and communities of practice, at different levels of proficiency, and situated in different disciplines, domains or job settings. Finding and choosing an appropriate CDP is not that easy for learners. The CDP has to match learners’ individual competence goals (e.g., MSc in Psychology), and their personal preferences (e.g., study location, didactical methods). Also, the course entry requirements should match learners’ already acquired competence profile (e.g., bachelor degree in psychology). Many learners do not have the adequate skills to find out which CDPs are offered and which are appropriate. Thus, these learners need to be supported when looking for an appropriate CDP. Within educational institutes, study advisers can be consulted for this, but when choices concern learning activities of more than one educational institute or even outside institutes, advice is scarce. Recently, Personal Recommender Systems (PRS) for learners are developed for that purpose. A PRS provides personal recommendations for learners aimed at finding and selecting CDPs that best match their needs (Drachsler, Hummel, & Koper, in press; Van Setten, 2006; Koper, 2006).
In an information-based PRS for learners, information concerning desired and acquired competence profiles are combined (Hummel et al., 2007).

In order for a PRS to provide personal recommendations for a learner, a competence description is needed that enables comparison of information concerning individual target competences and acquired competences on the one hand, and CDP-related information on the other. In this article, first the current competence descriptions will be described and examined for their usefulness for PRSs. Second, the term context will be conceptualized and the term learning and performance situation (LP-situation) will be introduced. It is argued that relevant complexity factors typify a LP-situation. Third, advantages and disadvantages of including LP-situations and, consequently, complexity factors, in characterizing CDPs are addressed. We claim that, although valuable for its purpose, current competence descriptions should be extended with a description of the LP-situation. An adjusted, augmented competence description, including LP-situations, is suggested. Fourth, an example of a CDP selection problem is presented that demonstrates how PRSs can benefit from an augmented competence description. In the last section of this article, implications of our claim will be discussed.

Current competence descriptions

The concept of competence can have quite different connotations and definitions (Cheetham & Chivers, 2005; Stoof et al., 2002; Van Merriënboer, Van der Klink, & Hendriks, 2002; Westera, 2001). It should also be noted that there is a distinction in the literature between the term ‘competence’ and the term ‘competency’ (De Coi et al., 2006; Eraut, 1994). Competence is given a generic or holistic meaning and refers to a person’s overall capacity whereas competency refers to specific capabilities (knowledge, skill, attitude, ability). Cheetham and Chivers (2005) offer the following rather general definition of competence:

Effective overall performance within an occupation, which may range from the basic level of proficiency through the highest levels of excellence.

Stoof et al. (2002), on the other hand, postulate that the meaning of the concept of competence is very unclear. They give a short overview of recent history of ‘competence’ and provide examples of current definitions, such as “a cluster of knowledge, skills and attitudes” or “the ability to handle a situation”. Stoof and colleagues conclude that it is useless to look for the true definition of competence and argue that everyone may construct their own competence definition instead, as long as it is viable. Viability of a competence definition increases when it is clear what the representations and opinions about competences are of the people who construct the competence definition. In addition, the goal of the competence definition should be made clear in order to construct a suitable and useful definition. Finally, it should be clear who the intended users of the definition are (Stoof et al., 2002).

However, idiosyncratic definitions of competence are insufficient for enabling system-based personal recommendations for selecting adequate CDPs. These recommendations could be based on learners’ needs (i.e., their competence goals), their preferences (e.g., preferred study mode, preferred learning style, preferred delivery mode, preferred task characteristics such as performance situation), and CDP-related information. Thus, for personal recommendations, retrieval, exchange and reuse of learning units for international educational institutes is needed. A learning unit refers to each unit where learning can take place, and it can be large or small. Examples are a course, a module, and a CDP. For an effective exchange of learning units, educational institutes need to use a common format of competence description. In the same vein, a common format of competence description is needed when educational designers aim to design formal CDPs that could be used and reused by international educational institutes. These designers of CDPs, as well as the users of the programs, need to know what learners should be able to do when learners have completed a CDP, that is, which competences should be acquired in the CDP. Thus, designers should make sure that they explicitly describe the necessary elements of the competence aimed at in the designed CDPs. Moreover, learners want to know what competences are needed for a particular job (the so-called job profile or required competence profile), what competences they already have acquired (their acquired competence profile, e.g., accreditation of prior learning), what competences still have to be acquired (their competence gap profile), and where to find existing CDPs to reduce the gap between the acquired competence profile and job profile. For the goals of learners, educational designers, and educational institutes, a sound competence description or model that specifies all relevant ingredients is needed.
Some valuable initiatives on standardization of modelling competencies exist, such as those of IMS RDCEO (2002), IEEE-RCD (2006), and HR-XML (2006). The main purpose of these initiatives is to enable interoperability among learning systems that deal with competency information by providing a means for them to refer to common definitions with common meanings. Central repositories are build that define competencies and these competency definitions can be referenced by external data structures. All three definitions include titles and descriptions that need to be interpreted by human beings. Furthermore, the objective of these descriptions is to represent formally the key characteristics of a competency, independently of its use in any particular context or environment. Thus, these approaches to modelling competencies exclude ‘context’ from their definitions, because when information concerning context becomes part of the competency definition, its reusability is drastically reduced (De Coi et al., 2006). On the other hand, when selecting an adequate CDP, the context to which a CDP refers to may be very important to the learner. For instance, a professional teacher who wants to develop her teaching competences may particularly look for urban, cross-cultural work situations. Thus, for adequate recommendations, PRSs should be able to retrieve and exchange information concerning context. Several theorists (e.g., Sandberg, 2000) argue that competences used in accomplishing work are not primarily context-free but are situational, or context dependent. Also Koper (2006), in his definition of competence, links competence to context or situations, by him labelled as ‘ecological niche’ (an occupation, a hobby, a market, a sport, etc.). We conclude that context is an important element related to competence and that context should be modelled. In order to maximize reuse, competence and context should be considered as different dimensions that should be modelled separately (De Coi et al., 2006).

**Conceptualizing context**

Tessmer and Richey (1997, p. 87) define context as “multilevel body of factors in which learning and performance are embedded”. Context is thus perceived as the simultaneous interaction of a number of mutually influential factors. The multi-level nature of context means that different spatial and temporal levels of contexts need to be considered, such as the immediate and surrounding contexts.

According to Cheetham and Chivers (2005), a person could be extremely competent in one particular context, but becomes much less so if the context or environment changes. For instance, a teacher in primary education can feel highly confident in his ability to control a class of a rural town school, but a definite lack of confidence when supposed to control a class of a school in a big city. Here, both contexts ask for similar competences (e.g., ‘interpersonal competence’), but one would agree that the latter ‘context’ is much more complex. This difference in complexity is determined by several complexity factors, such as class size or social economic background of learners and parents. In the revised competence model of Cheetham and Chivers (2005), context of work has an important place. They define context of work quite generally as ‘the particular working situation in which an individual is required to operate’. Also Van Merriënboer, Van der Klink and Hendriks (2002) argue that competences are context-specific. They examined several competence definitions, representing eleven approaches to competences. Nearly all approaches that were mentioned in the study of Van Merriënboer et al. emphasized the context-specificity. That doesn’t mean, however, that all theorists perceived ‘context’ the same way. As a matter of fact, Van Merriënboer et al. concluded that ‘context’ is hardly defined at all. According to De Coi et al. (2006), modeling context may be a complex task, as it may coincide with a whole domain. So are we replacing the problem of defining competence by the problem of defining context? Not really. We argue that identifying the most relevant complexity factors to typify a ‘context’ could make the concept of ‘context’ valuable and usable for exchange and reuse.

Instead of ‘context’, we prefer the term *learning and performance situation* (LP-situation) for two reasons. First, ‘context’ can refer to very abstract or concrete notion of circumstances such as (1) a culture or environment (e.g., a school in a suburb), (2) types of situations (e.g., classes with medium class size and children of two cultures, with complexity factor class size set to medium and complexity factor cultural diversity set to two), and (3) to very specific cases (e.g., a particular class with John, Paul, George, Mohammed, Ahmed, …). It is the middle level of abstraction, that is, types of situations that we consider to be appropriate for reuse. Second, we like to speak of ‘learning and performance situation’ rather than of ‘work situation’ or just ‘situation’. In education as well as in professional development, the actions learners perform when acquiring competences can also refer to other situations then situations directly related to work or occupations whereas such actions can still be very important for the acquisition of relevant competences. These other situations may have a lot of characteristics in common with the work situations, though.
Advantages and disadvantages of identifying complexity factors

LP-situations are considered to be typified by interplay of various complexity factors. Including LP-situations and, consequently, complexity factors, in characterizing CDPs has advantages as well as disadvantages. The first, and most important advantage, is that by including LP-situations in characterizing CDPs, personal recommendations can be tuned to the needs of the learner and the learner will be confronted with most relevant tasks when actually participating in a CDP. Although educational institutions nowadays provide descriptions of competences and sometimes descriptions of critical or characteristic job situations, these descriptions of job situations are generally very shallow and lack a systematic approach such as we propose in our LP-situation concept. Second, LP-situation may support the determination of the proficiency levels of competences which are used within a competence specification/standard. Third, a series of LP-situations and the complexity factors within it, ranging from relatively easy to complex, may be very helpful for the design of learning tasks. Fourth, descriptions of relevant performance situations may be useful for accreditation of prior knowledge. As our approach makes it possible to have an unequivocal mapping from LP-situation to proficiency levels within a competence specification standard, one can argue that LP-situations comply with the need to be able to exchange learning units between CDPs. Consequently, any lifelong learner will benefit from a competence description using LP-situations.

A disadvantage of the inclusion of LP-situations in competence descriptions could be that it makes a competence description more complex. Moreover, agreement between stakeholders (learners, educational institutes, professions and even politicians) concerning the relevant LP-situations and the corresponding complexity factors should be established. This will probably impede the debate and negotiation with respect to competences and proficiency levels for domains, whereas already existing agreements on competence maps need to be extended with LP-situations.

We will illustrate our claims with an example. Imagine that a learner wants to learn how to ride an All Terrain Bike (ATB) in all circumstances. In Table 1, the already acquired competence profile, the target competence profile and the complexity factors of LP-situations are specified. The already acquired competence profile refers to the collection of a person’s already acquired competences that are relevant for the competence goal. The desired competence profile refers to the collection of relevant competences the person needs to achieve for proving he/she has reached the competence goal. LP-situations are created by combining several levels of relevant complexity factors. In this example, with four complexity factors that all have two levels, sixteen different LP-situations are distinguished.

---

Table 1. Use case riding an All Terrain Bike (ATB) for a learner with a specific goal and preferences

<table>
<thead>
<tr>
<th>Goal:</th>
<th>I want to be able to ride an ATB in all circumstances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquired competence profile:</td>
<td>I can ride an ATB on a paved, quiet, and flat place when the sun is shining</td>
</tr>
<tr>
<td>Desired competence profile</td>
<td>I can ride an ATB when descending an unpaved mountain highway during a traffic peak and heavy rainfall</td>
</tr>
<tr>
<td>Complexity factors:</td>
<td>a. surface structure of the road</td>
</tr>
<tr>
<td></td>
<td>b. amount of traffic on the road</td>
</tr>
<tr>
<td></td>
<td>c. shape of the road</td>
</tr>
<tr>
<td></td>
<td>d. weather conditions</td>
</tr>
<tr>
<td>Values for complexity factors:</td>
<td>a. surface structure of the road (paved, unpaved)</td>
</tr>
<tr>
<td></td>
<td>b. amount of traffic on the road (quiet, peak)</td>
</tr>
<tr>
<td></td>
<td>c. shape of the road (flat &amp; straight, hilly &amp; curved)</td>
</tr>
<tr>
<td></td>
<td>d. weather conditions (bright &amp; sunny, heavy rain)</td>
</tr>
<tr>
<td>Preferences for the learner:</td>
<td>I can maximally spend 20 hours a week on training</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, by considering LP-situations in characterizing CDPs, personal recommendations can be tuned to the needs of the learner (first advantage), and the learner will be confronted with most relevant tasks when actually participating in a CDP. In our example, a personal recommendation system will search for CDPs that – in the end - include cases (most specific level of abstraction) for riding an ATB on unpaved, hilly and curved roads, during traffic peaks and during heavy rainfall which will not exceed spending maximally 20 hours a week in training sessions. A case is considered as an instantiation of a LP-situation, including an assignment or task. At the level of concrete learning materials included in the CDP, this could be a case on descending the road safely just outside Adis Abeda at 8 o’clock in the morning during the rainy season, but it could also be a case on the Karakoram
LP-situations can also support the determination of the proficiency levels of competences within a competence specification/standard (second advantage). Educational institutes as well as work organizations use different idiosyncratic scales to represent proficiency level, but they are often arbitrary, because the grounds on which the standards for each proficiency level are determined are very often unclear (Eraut, 1994). We suggest using the complexity factors and their scales within LP-situations for the determination of proficiency levels. For instance, a learner could very well ride an ATB on a paved quiet road, but not on an unpaved road during peak hours. Similarly, a teacher could very well act effectively in a situation in which s/he has to instruct ten students without learning problems. However, the same teacher may fail to act adequately in a situation in which s/he has to instruct thirty students with learning problems. These two situations differ concerning the ‘number of students’ and the ‘number of students with learning problems’. These two complexity factors, in fact, determine the LP-situation. The values of these complexity factors in a specific situation could very well determine the proficiency level on which a teacher can act effectively. By combining the complexity factors, one could create a scale of several levels of proficiency. For instance, a combination of two complexity levels of complexity factor A (values a and A) and two complexity levels of complexity factor B (values b and B) results in four different complexity levels (ab, Ab, aB, and AB), creating a scale with three or four proficiency levels. When several complexity levels can be identified, as is the case in job situations of teachers, and when these factors have more than two levels, the number of proficiency levels may increase very fast. It is therefore suggested not to just determine and rank proficiency levels but also to explicitly describe proficiency levels in terms of complexity factors. This issue will be addressed when a worked out example is presented at the end of this article.

Third, performance complexity characterized via complexity factors may be helpful for the design of learning tasks and CDPs (third advantage). In his 4C/ID-model, Van Merriënboer (1997), advocates a whole-task approach of instruction. In a whole-task approach, the learner is taught all constituent skills at the same time, but conditions under which the whole skill is trained become more complex during the training. Conditions that may simplify the performance of a complex skill, that is, complexity factors, must be identified in order to create authentic cases that differ in complexity. A learner starts with the simplest authentic case that a professional may encounter. During the training, the simplifying conditions should be relaxed one at a time, so that the cases for instruction become more and more complex. For instance, suppose that a novice teacher in training wants to work on her ‘competence for collaboration with the working environment’. For this, she must, among other things, keep in touch with students’ parents or guardians, give them professional information, and use the information she gets from them. Authentic cases that differ in complexity must be created, for instance cases in which the teacher in training has a meeting with student’s parents. Complexity factors for such an authentic case could be: (1) the possibility of learning problems of the student, (2) the possibility of social problems of the students, (3) the social skills of the parents, and (4) whether the student joins the meeting between teacher and parents. In our view, a novice teacher in training should start practicing with an authentic case in which she meets socially skilled parents of a student without learning and social problems, with the student being absent. Thus, all complexity factors should be set on the simplest option. In that relatively simple authentic case, the teacher in training can practice all the skills, procedures and scripts that are relevant for adequate performance. During training, the complexity factors can be set to a more complex value one at the time. It should be noticed that for every case, no matter how complex, learners should meet similar performance criteria. For instance, in all cases concerning meetings with parents or guardians, the teacher should be able to (1) explicit the goal of the meeting, (2) provide relevant information to the parents, (3) get relevant information from the parents, (4) sustain a good relationship with the parents, and (5) make appointments for future activities. At the Open University of The Netherlands, a large set of job situations and learning tasks for teachers were designed based on variation of complexity factors (Proformas, 2008).

Fourth, it is increasingly acknowledged that part of the learning does not take place in formal but in non-formal and informal situations. When enrolling formal education, for instance Psychology at the Open University of the Netherlands, learners may already have acquired competences in informal situations that are also part of the formal CDP (i.e., curriculum offered by the Open University of the Netherlands). But how could accreditation of prior learning be reached? What information should the learners provide in order to convince the Psychology institute of
the fact that the learner has already acquired relevant competences? We suggest that the LP-situation, mapped towards a competence specification/standard, may function as the missing link between informal learning and formal accreditation (fourth advantage). For instance, an individual who has been a volunteer counselor for Kids Help Phone received an informal training and has a lot of experience with counseling conversations. The instantiations of the LP-situations in which the individual acted effectively, that is, particular cases, show many similarities with the LP-situations in which a clinical psychologist could be involved, including roles and performance indicators. When this is the case, the learners may be considered for exemption for some modules of the clinical psychology curriculum.

Thus, in our opinion, professionals and educational institutes should describe their LP-situations when characterizing their CDPs. It is interesting to notice that some initiatives are aimed at describing characteristics of situations in which professionals are supposed to perform. For instance, the Occupational Information Network (O*Net, http://online.onetcenter.org/) is a comprehensive database of worker attributes and job characteristics. O*NET is being developed as a timely, easy-to-use resource that supports public and private sector efforts to identify and develop the skills of the American workforce. It provides a common language for defining and describing occupations. Its flexible design also captures rapidly changing job requirements. Part of O*Net is a description of work context. Work context of occupations is determined by a 57 item questionnaire in which several dimensions of work context are listed, for instance, contact with others, responsibility for health and safety, conflict situations, and telephone conversations. Many of these items refer to work conditions (temperature, body vibration, radiation) but some of the 57 items can be used for a description of a LP-situation, as long as it concerns a relevant complexity factor for the particular job or occupation. Note that a complexity factor must be a variable that, depending on its value, makes performance for a professional in his or her job, more simple or complex.

Thus, in the proposed competence description, competence is linked to proficiency level, which, in turn, is determined by complexity factors of several LP-situations. Evidence of acquired competences will be based on performance in instantiated LP-situations, that is, in cases of particular complexity. In our view, a competence description that is useful for a PRS contains the elements that are specified in Table 2.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Explanation</th>
<th>Reqd</th>
<th>Mult</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Competence description</td>
<td>This element specifies the competence description.</td>
<td>-</td>
<td>-</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.1</td>
<td>Identifier</td>
<td>A unique label that identifies this competence description (ID).</td>
<td>M</td>
<td>1</td>
<td>ID</td>
</tr>
<tr>
<td>0.2</td>
<td>Title</td>
<td>A single text label for the competence description. This is a human-readable name for the competence. The title may be repeated in multiple languages.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.3</td>
<td>Description</td>
<td>A human-readable description of the competence. Unstructured string meant to be interpretable only for humans. The description may be repeated in multiple languages.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.4</td>
<td>Definition</td>
<td>A structured definition of the competence description.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.5</td>
<td>Learning-Performance-Situation-ref</td>
<td>Refers to a learning performance situation.</td>
<td>O</td>
<td>0..*</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.6</td>
<td>Complexity-Factor-ref</td>
<td>Refers to a complexity factor.</td>
<td>M</td>
<td>1..*</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.7</td>
<td>Proficiency-Level-ref</td>
<td>Refers to a proficiency level.</td>
<td>M</td>
<td>1</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.8</td>
<td>Performance-Indicator-ref</td>
<td>Refers to a performance indicator.</td>
<td>O</td>
<td>0..1</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.9</td>
<td>Mapping-function</td>
<td>Description of how the multiple values of the complexity variables are mapped towards all single-value proficiency levels. All possible combinations should be mapped towards a proficiency level and all possible values for proficiency level should be used at least once.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.10</td>
<td>Metadata</td>
<td>Placeholder for metadata. Include IMS Meta-Data here, using its namespace.</td>
<td>O</td>
<td>0..1</td>
<td>Sequence</td>
</tr>
</tbody>
</table>
### Learning-Performance-Situation

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Explanation</th>
<th>Req</th>
<th>Mult</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Learning-Performance-Situation</td>
<td>This element contains a sequence of elements for learning and performance situations definitions.</td>
<td>O</td>
<td>0..*</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.1</td>
<td>Identifier</td>
<td>A unique label that identifies this learning performance situation.</td>
<td>M</td>
<td>1</td>
<td>ID</td>
</tr>
<tr>
<td>0.2</td>
<td>Title</td>
<td>A single text label for the learning and performance situation. This is a human-readable name for the learning and performance situations. The title may be repeated in multiple languages.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.3</td>
<td>Description</td>
<td>A human-readable description of the learning and performance situation. Unstructured string meant to be interpretable only for humans. The description may be repeated in multiple languages.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.4</td>
<td>Definition</td>
<td>A structured definition of the learning and performance situations.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
</tbody>
</table>

### Complexity-Factor

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Explanation</th>
<th>Req</th>
<th>Mult</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Complexity-Factor</td>
<td>This element contains a sequence of elements for complexity factor definitions.</td>
<td>M</td>
<td>1..*</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.1</td>
<td>Identifier</td>
<td>A globally unique label that identifies the complexity factor.</td>
<td>M</td>
<td>1</td>
<td>ID</td>
</tr>
<tr>
<td>0.2</td>
<td>Title</td>
<td>Text label of the complexity factor that has impact on the complexity of a learning and performance situation.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.3</td>
<td>Description</td>
<td>Description of the complexity factor that has impact on the complexity of a learning and performance situation.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.4</td>
<td>Definition</td>
<td>A structured definition of the complexity factor.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.5</td>
<td>{itemvalue}</td>
<td>Schema group that enumerates the possible values of each complexity factor.</td>
<td>M</td>
<td>1</td>
<td>Group</td>
</tr>
</tbody>
</table>

### Proficiency-Level

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Explanation</th>
<th>Req</th>
<th>Mult</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Proficiency-Level</td>
<td>This element contains a sequence of elements for proficiency level definitions.</td>
<td>M</td>
<td>1</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.1</td>
<td>Identifier</td>
<td>A globally unique label that identifies the proficiency level.</td>
<td>M</td>
<td>1</td>
<td>ID</td>
</tr>
<tr>
<td>0.2</td>
<td>Title</td>
<td>Text label of the proficiency level.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.3</td>
<td>Description</td>
<td>A human-readable description of the proficiency level.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.4</td>
<td>Definition</td>
<td>A structured definition of the proficiency level.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.5</td>
<td>{itemvalue}</td>
<td>Schema group that enumerates the values of the proficiency levels.</td>
<td>M</td>
<td>1</td>
<td>Group</td>
</tr>
</tbody>
</table>

### Performance-Indicator

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Explanation</th>
<th>Req</th>
<th>Mult</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Performance-Indicator</td>
<td>This element contains a sequence of elements for performance indicator definitions.</td>
<td>O</td>
<td>0..1</td>
<td>Sequence</td>
</tr>
<tr>
<td>0.1</td>
<td>Identifier</td>
<td>A globally unique label that identifies the performance indicator.</td>
<td>M</td>
<td>1</td>
<td>ID</td>
</tr>
<tr>
<td>0.2</td>
<td>Title</td>
<td>Text label of the performance indicator.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.3</td>
<td>Description</td>
<td>A human-readable description of the performance indicator.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.4</td>
<td>Definition</td>
<td>A structured definition of the performance indicator.</td>
<td>O</td>
<td>0..1</td>
<td>String</td>
</tr>
<tr>
<td>0.5</td>
<td>{itemvalue}</td>
<td>Schema group that enumerates the values of the performance indicators.</td>
<td>O</td>
<td>1</td>
<td>Group</td>
</tr>
</tbody>
</table>

### {itemvalue}

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Explanation</th>
<th>Req</th>
<th>Mult</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Itemvalue</td>
<td>A schema group for values.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>Identifier</td>
<td>A globally unique label that identifies the item.</td>
<td>M</td>
<td>1</td>
<td>ID</td>
</tr>
</tbody>
</table>
An example of a CDP selection problem

To illustrate the claims above, in this paragraph an example of a CDP selection problem is presented. The example is based on the competence requirements for teachers, defined by the Association for the Professional Quality of Teachers, SBL (SBL, 2004). Imagine that a young teacher wants to acquire the competence for collaboration with the working environment of a teacher in pre-higher education. There are many CDPs around in several educational institutes for attaining this competence goal. The young teacher decides to consult a PRS. Besides using personal preferences within PRS, additional input for the PRS is the already acquired competence profile, consisting of relevant certificates as well as experiences in relevant job situations, and the desired competence profile. The experiences in relevant job situations may very well be matched with LP-situations with several values of corresponding complexity factors. Which of the LP-situations is the young teacher familiar with? In which of the LP-situations, including values of corresponding complexity factors, was the young teacher successful? Similarly, the young teacher can identify in which of the LP-situations, including values of corresponding complexity factors, wants the young teacher to be successful. It is probably exceptional when the young teacher knows exactly what his desired competence profile is, but also incomplete information about the desired competence profile can put the PRS to work. It is possible when the PRS provides recommendations, the desired competence profile may become clearer to the young teacher. In our example, the young teacher wants to perform adequately in the LP-situation planned meetings with parents of students with social problems.

After the young teacher’s input of information concerning acquired and desired competence profile, as well as personal preferences, the PRS will search for CDPs that match personal needs, preferences, and competence profiles. Subsequently, the PRS recommends and provides access to possible cases, that is, possible learning activities in instantiated LP-situations with determined values of complexity factors, available at different educational institutes. In order to do so, a competence description is needed in which LP-situations, with corresponding complexity factors, are identified. In Table 3, an overview of a possible competence description of this example is presented, including proficiency levels, performance situation, complexity factors and values. In this example, the possible itemvalues of the complexity factor “social problems of the student” are 0 (student has no social problems) and 1 (student has social problems). The combination of itemvalues of the relevant complexity factors determines the difficulty of LP-situations. In our example, there are four relevant complexity factors (i.e., the possibility of learning problems of the student, the possibility of social problems of the students, the social skills of the parents, and whether the student joins the meeting between teacher and parents). The second complexity factor is described in Table 3. Furthermore, the combination of itemvalues of the four relevant complexity factors requires a mapping towards the allowed proficiency levels. For instance, when a learner cannot perform adequately in a LP-situation in which the values of the four relevant complexity factors are 0, the corresponding proficiency level may become 1 (novice). When a learner performs adequately in a LP-situation in which the values of the four relevant complexity factors are 0, but not adequately in a LP-situation in which at least one of the values of the complexity factors is 1, the corresponding value of the proficiency level may become 2 (advanced beginner), as is illustrated in the example. Similarly, when the learner performs adequately in a LP-situation in which the values of the four relevant complexity factors are 1, the proficiency level may become 5 (expert). More intermediate levels can be determined accordingly. The higher the number of complexity factors or the higher the number of levels of a complexity factor, the more proficiency levels are possible and the more important it is to have a good mapping between values of complexity factors and proficiency levels. For instance, suppose that in a LP-situation complexity factor A has possible values (a1, a2), complexity factor B has values (b1, b2) and complexity factor C has values (c1, c2). All possible combinations (a1, b1, c1), (a1, b1, c2), (a1, b2, c1), (a1, b2, c2), (a2, b1, c1), (a2, b1, c2) at the one side and proficiency levels at the other should be mapped. When there are less than nine proficiency levels allowed (and we believe that nine proficiency levels are too many), it should be made clear how these nine combinations are mapped
towards a limited set of proficiency levels (four in our example). Thus, the determination of a limited set of proficiency levels should be based on an explicit rule or description.

In Table 3, one LP-situation is described in terms of complexity factors and performance indicators. The example aims at illustrating our claims and is thus not a complete description of this competence. That means that other LP-situations as well as complexity factors can be added to the example. One competence can refer to many LP-situations, and many complexity factors may determine a LP-situation. Therefore, it is important to reduce the number of LP-situations and the number of complexity factors.

The identifiers and itemvalues of proficiency levels and the complexity factors corresponding to a LP-situation, can be input for PRSs. When the young teacher in our example wants to find cases at proficiency level 2 (advanced beginner) to develop the competence for collaboration with the working environment of a teacher in pre-higher education, the PRS system selects cases (instantiations of LP-situations) corresponding to proficiency level 2, that is, it selects instantiations of LP-situations of which, for instance, only one complexity factor has the value 1 and three have the value 0. The young teacher can be considered to be at proficiency level 2 when adequate performance is shown in cases corresponding to LP-situations of proficiency level 2.

This example illustrated the first, second and fourth advantage of matching LP-situation to a competence description. In order for the PRS to select appropriate CDPs based on LP-situations and complexity factors, learning activities (cases) and CDPs should be designed based on a variation of these LP-situations. These LP-situations and complexity factors are helpful for the purposes of competence descriptions that we described above.

### Table 3. Example of a competence description

<table>
<thead>
<tr>
<th>Name</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 CompetenceDescription</td>
<td></td>
</tr>
<tr>
<td>0.1 Identifier</td>
<td></td>
</tr>
<tr>
<td>0.2 Title</td>
<td>Competence for collaboration with the working environment</td>
</tr>
<tr>
<td>0.3 Description</td>
<td>The teacher in pre-higher education must keep in touch with the students’ parents or guardians, and with colleagues of educational and youth welfare institutions his/her school collaborates with. He/she must make sure that his/her professional actions are in line with those of others outside the school. Furthermore, he/she must contribute to a good development of collaboration between his/her school and the institutions concerned.</td>
</tr>
<tr>
<td>0.4 Definition</td>
<td>The teacher in pre-higher education must keep in touch with the students’ parents or guardians, and with colleagues of educational and youth welfare institutions his/her school collaborates with.</td>
</tr>
<tr>
<td>0.5 Learning-Performance-Situation</td>
<td></td>
</tr>
<tr>
<td>0.5.1 Title</td>
<td>Planned meeting with parents or guardians</td>
</tr>
<tr>
<td>0.5.2 Description</td>
<td>In a planned meeting, a teacher meets with student’s parents or guardians to discuss cognitive, social, and/or affective progress of the student. The teacher gives parents and other parties involved professional information about the students, and uses information the teacher gets from them.</td>
</tr>
<tr>
<td>0.5.3 Definition</td>
<td>In a planned meeting, a teacher meets with student’s parents or guardians to discuss cognitive, social, and/or affective progress of the student.</td>
</tr>
<tr>
<td>0.6 Complexity-Factor</td>
<td></td>
</tr>
<tr>
<td>0.6.1 Identifier</td>
<td></td>
</tr>
<tr>
<td>0.6.2 Title</td>
<td>Social problems of student</td>
</tr>
<tr>
<td>0.6.3 Description</td>
<td>The student has social problems, such as aggressive behavior in school, …</td>
</tr>
<tr>
<td>0.6.4 Definition</td>
<td>The student has social problems.</td>
</tr>
</tbody>
</table>

149
Discussion and conclusion

In this article, it was argued that a clear-cut description of the concept of ‘competence’ to characterize CDPs, individuals’ acquired and target competence profile, is needed for system-based personal recommendations for selecting appropriate CDPs. We proposed to extend current initiatives on standardization of modelling competencies with the concept of ‘LP-situation’ and claimed that this extension has added value for personal recommendations for adequate CDP selection, for determining proficiency levels of competences, for the design of learning tasks, and for accreditation of prior knowledge. With an adequate CDP, at the right proficiency level, that is, consisting of LP-situations with values for complexity factors that match learners’ prior knowledge and desires, learners can acquire competences in a more efficient, less frustrating way. Therefore, competence descriptions, CDPs, LP-situations with values for complexity factors and proficiency levels for competences should be clearly defined. Moreover, an educational technological system such as a PRS should be able to deal with the extensive information concerned with competence descriptions, CDPs, and LP-situations.

This leads to the question how and by whom the LP-situations should be designed. We suggest that professionals and educational institutes should collaboratively describe LP-situations for their domain in a systematic way, including complexity factors and their possible values, and, for each competence agree on the mapping towards proficiency levels. In the Netherlands, the Association for the Professional Quality of Teachers (SBL, 2004) succeeded in agreeing on the competence requirements for teachers. The next step is to agree on a systematic description of LP-situations. Further research is needed to examine whether professions are willing and able to add LP-situations and corresponding complexity factors to job descriptions and competence requirements. Moreover, research is needed to determine methods and procedures for communities of practitioners to define characteristic LP-situation with corresponding complexity factors. Special tools, such as the web-based support for constructing competence maps of Stoof et al. (2007), could be designed or adjusted for explicit support concerning systematic description of LP-situations.

Not all competences are equally context-specific (Van Merriënboer et al., 2002). Some competences are applicable to many LP-situations of many domains (e.g., social competence) and others are limited to a specific domain or to a few characteristic LP-situations. Consequently, general competences can be linked to many LP-situations, and thus, learners almost have an unlimited choice of LP-situations that can be used to develop this general competence. On the other hand, many relevant but domain-independent LP-situations could make it more difficult to distinguish a limited number of proficiency levels. It is a challenge for educational designers to select the appropriate LP-situations in a particular domain. For practical reasons, the number of relevant complexity factors and their possible values should be limited; otherwise many proficiency levels could be distinguished. It is open to discussion how many complexity factors and how many values for each of them are relevant to distinguish. Suppose one already agreed to distinguish between, lets say, three proficiency levels, then one needs to specify how the complexity factors and their possible values will be mapped towards those three proficiency levels. If the acquired competence profile would equal level one and desired competence profile would equal level three, personal recommendations would firstly offer CDPs that aim at level two. Often, the model of Dreyfus (2000) is used, in which five proficiency...
levels are described. In our opinion, each of these levels should be described in terms of LP-situations, which makes the proficiency levels concrete and attainable.

After agreement about the LP-situations and the complexity factors in the community of practitioners, technology should come in to define the competences in terms of the current initiatives such as RDCEO, HR-XML, and IEEE RCD and link these definitions to modeled LP-situations. In fact, we propose to add extensions for LP-situations to these initiatives in a way similar as is suggested by Sampson, Karampiperis, and Fytros (2007). In this way, our approach towards competence description complies with existing initiatives for referencing and exchanging competences between learning systems, human resource systems, and competency or skill repositories (De Coi et al., 2006).

Acknowledgement

The work on this paper has been sponsored by the TENCompetence Integrated Project that is funded by the European Commission’s 6th Framework Programme, priority IST/Technology Enhanced Learning (IST-2004-02787). (http://www.tencompetence.org)

References


Personalized Intelligent Mobile Learning System for Supporting Effective English Learning

Chih-Ming Chen
Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taiwan // chencm@nccu.edu.tw

Shih-Hsun Hsu
Graduate Institute of Network Learning Technology, National Central University, Taiwan // jason@cl.ncu.edu.tw

ABSTRACT
Since English has been an international language, how to enhance English levels of people by useful computer assisted learning forms or tools is a critical issue in non-English speaking countries because it definitely affects the overall competition ability of a country. With the rapid growth of wireless and mobile technologies, the mobile learning has been gradually considered as a novel and effective learning form because it inherits all the advantages of e-learning as well as breaks the limitations of learning time and space occurring in the traditional classroom learning. To provide an effective and flexible learning environment for English learning, this study adopts the advantages of the mobile learning to present a personalized intelligent mobile learning system (PIMS) which can appropriately recommend English news articles to learners based on the learners’ reading abilities evaluated by the proposed fuzzy Item Response Theory (FIRT). In addition, to promote the reading abilities of English news, the unknown or unfamiliar vocabularies of individual learner can also be automatically discovered and retrieved from the reading English news articles by the PIMS system according to the English vocabulary ability of individual learner for enhancing vocabulary learning. Currently, the PIMS system has been successfully implemented on the personal digital assistant (PDA) to provide personalized mobile learning for promoting the reading ability of English news. Experimental results indicated that the proposed system provides an efficient and effective mobile learning mechanism by adaptively recommending English news articles as well as enhancing unknown or unfamiliar vocabularies’ learning for individual learners.

Keywords
Mobile learning, Personalization, Intelligent tutoring system, English Learning

Introduction
Traditional teaching resources, such as textbooks, typically guide the learners to follow fixed sequence to other subject-related sections related to the current one during learning processes. Web-based instruction researchers have given considerable attention to flexible curriculum sequencing control to provide adaptable, personalized learning programs; therefore, many powerful personalized/adaptive guidance mechanisms, such as adaptive presentation, adaptive navigation support, curriculum sequencing, and intelligent analysis of learner’s solutions, were proposed to aid more efficient learning (Lee, 2001; Tang & Mccalla, 2003; Papanikolaou & Grigoriadou, 2002). Currently, most adaptive/personalized tutoring systems consider learner/user preferences, interests, and browsing behavior when investigating learner behavior for personalized services. However, these systems neglect the importance of learner ability when implementing personalized mechanisms. On the other hand, some researchers emphasized that personalization should consider levels of learner knowledge, especially in relation to learning (Chen et al., 2005; Chen et al., 2006; Brusilovsky, 1999). That is, considering learner ability based on major fields and subjects can promote personalized learning performance.

In recent years, the fastest growing web community is mobile visitors who browse web pages or retrieval web information using PDAs or cell phones by wireless networks. Similarly, the development of educational technologies also tends to be more mobilized, portable, and personalized. These trends lead to the learning form is dramatically changing from the traditional classroom learning to the electronic learning (E-learning) (Lin & Hsieh, 2001), mobile learning (M-learning) (Chang et al., 2003; Cabrera et al., 2005; Ting, 2005; Holzinger et al., 2005) or even ubiquitous learning (U-learning) (Wang, 2004; Rogers et al., 2005; Tummala & Jones, 2005; Wilkerson et al., 2005) due to fast development of communication technology. Among these novel learning forms, mobile learning
has been considered as an effective form of flexible learning because it is possible to break the limitation of teaching place where must take place in a classroom or located computers. In the meanwhile, mobile learning is also helpful to perform learning activities utilizing learner’s spare time at any time from any place with wireless networks.

Since English has been the most important second language (L2) in many non-English speaking countries, developing useful computer assisted learning forms or tools for supporting effective English learning is a critical issue in the English-language education field (Collins, 2005; Shih, 2005). Learning English involves memorization and practice of a large number of vocabulary words and grammatical structures. Recently, some scenarios of mobile learning have been successfully proposed to aid language learning activities of outside classroom, such as TenselITS, a mobile intelligent tutoring system with learner’s location awareness for supporting language learning, designed primarily for Chinese learners of English (Cui & Bull, 2005), cell phone assisted language learning system (Kiernan & Aizawa, 2004; Chinnery, 2006), and so on. Any language learning can be divided into four issues including listening, reading, speaking, and writing skills; however, Huckin et al. (1993) indicated that reading ability and vocabulary knowledge are two of the most important components of performance in a second language and depend on the other, especially in academic settings. In other words, reading can help language learners to acquire vocabulary and vocabulary knowledge is helpful to promote reading comprehension. Alderson (1984) proposed that the interpretation of words and syntactic structures - that is, grammar and vocabulary – is the main factor in poorer reading performance in the second language than in the first language (L1). Nation (1993) also proposed that the growth of reading skills is beneficial to develop knowledge of the world, and simultaneously increase basic academic and technical vocabularies. Moreover, Coady’s study (1997) pointed out that extensive reading is an effective way to promote learners’ reading performance. Generally, the extensive reading method frequently recommended by English teaching experts is to read storybooks, magazines, periodicals and newspapers printed in English. Hwang and Nation’s study (1989) emphasized that newspapers constantly repeat some frequent words and technological terms related to a topic. Multiple repeated vocabularies can reduce readers’ lexical loading and activate their imagination, experience, and background knowledge to better extract information from the reading text. Additionally, newspapers provide benefits to learn practical vocabularies frequently used in daily lives. However, readers often suffer many difficulties while reading English newspapers due to too many unfamiliar or unknown words used in various professional domains.

In addition, Jolly (1978) claimed that success in reading a foreign language mainly depends on one’s first-language reading ability rather than on the learners’ level of first-language. Namely, the foreign language reading is a reading problem, not a language problem. Coady (1973) thought that the reading process is much the same for all language. That is, good native language readers can be logically expected to be good second language readers. Yorio (1971) proposed that the reading problems of foreign language learners are due largely to imperfect knowledge of the language, and to native language interference in the reading process. However, some researches (Kern, 1994; Upton & Lee-Thompson, 2001) took a contrary view on the topic of “first language interference” to indicate that L2 readers often access to their L1 and use this resource as a strategy to help them comprehend an L2 text. Moreover, many empirical evidences (Alderson, 1984) showed that poor foreign language reading is due to incorrect strategies for reading that foreign language, strategies which differ from the strategies for reading the native language. Restated, poor foreign language reading is due to reading strategies in the first language not being employed in the foreign language. Furthermore, some researches (Carrell, 1991; Lee & Schallert, 1997) clearly mentioned that L2 proficiency influences overall L2 reading ability.

To conquer the problems of English learning, specially for reading and vocabulary acquisition, the common advice from many English teachers on the best way for learning English is “little and often” (Buda, 1984). Restated, a few minutes every day is certainly much better than a few hours once a month. Therefore, this study adopts the advantages of the mobile learning in terms of easily breaking the limitations of time and space and utilizing spare time for ubiquitous learning. Meanwhile, developing a personalized reading learning strategy to enhance reading learning of individual learners was also a concerned issue in the study. To integrate the mobile learning with the proposed personalized reading learning strategy, this study presents a personalized intelligent mobile learning system (PIMS) based on the proposed fuzzy Item Response Theory, which can conduct personalized curriculum sequencing, for supporting effective English reading learning for individual learners. Here, the electronic English news articles automatically retrieved from English news sites (FTV http://englishnews.ftv.com.tw/index.asp) by an intelligent crawler agent are used as the course materials for personalized mobile learning in the proposed PIMS system. Meanwhile, the unknown or unfamiliar vocabularies of individual learner can also be automatically discovered and retrieved from the reading English news articles by the PIMS system according to the English vocabulary ability of
individual learner for promoting reading learning by enhancing vocabulary learning. Experimental results indicated that the proposed PIMS system provides benefits to promote learner’s reading ability of English news articles, vocabulary ability, and learning interests due to providing personalized and flexible mobile learning mechanisms for individual learners.

System Design

This section describes the details of the proposed system architecture. An overview of system architecture is presented in the first subsection. The other subsections describe the English e-news archive procedures, detail the scheme of measuring difficulty of English news, explain the schemes of the personalized English news recommendation and personalized vocabulary recommendation.

System Architecture

A personalized intelligent mobile learning system (PIMS) supported by the personalized vocabulary learning system (Chen & Chung, 2006), which includes a remote courseware server, client mobile learning system, and data synchronized agent, is presented herein. Figure 1 shows the detailed system architecture. The client mobile learning system which consists of four intelligent agents and four databases can appropriately recommend English news articles for individual learners to enhance their reading abilities as well as vocabulary abilities for individual learners based on the proposed fuzzy Item Response Theory (FIRT). The remote courseware server containing three intelligent agents and one database aims at automatically collecting English news articles from the Internet by an intelligent crawler for the remote courseware & user portfolio database and evaluating the difficulty parameters of English news articles by the proposed scheme of measuring difficulty of English news articles. Moreover, to support the off-line learning mode, the data synchronized agent is in charge of keeping data consistency between the client databases with the server databases after the wireless network recovers on-line connection. In this work, the merge replication technique provided in Microsoft SQL server was employed to perform this work. The detailed system components are explained in the following subsections.

Figure 1. The system architecture of the proposed PIMS system

The Remote Courseware Server

In Fig. 1, the right part reveals the system architecture of the remote courseware server, which includes the English news crawler agent, difficulty assessment agent of English news, courseware management agent, and remote
Currently, many web sites have provided a large amount of free English news articles with corresponding Chinese translations for readers’ browsing in Taiwan, such as FTV and CTV English news sites (http://www.cts.com.tw/news/insidetaiwan/). These corresponding Chinese translations are helpful to speed up the comprehension degree of reading for Chinese readers while reading an English news article. To construct abundant English news courseware for the reading learning services of English news articles, the needed course materials are automatically retrieved and stored into the remote courseware & user portfolio database from the Internet by the English news crawler agent. The English news crawler agent aims at automatically gathering English news articles and extracting metadata contained in news articles from English e-news site to conduct the job of English E-news archive. Moreover, the difficulty assessment agent of English news is in charge of automatically measuring the difficulty parameters of English news articles according to vocabulary grading levels of General English Proficiency Test (http://www.gept.org.tw/indexie.asp) in Taiwan and the proposed reading ease formula modified from Flesch reading ease formula (Flesch, 1948). These measured difficulty parameters will be utilized to recommend appropriate English news articles to individual learners based on the proposed fuzzy Item Response Theory. Finally, the courseware management agent with authorized account management mechanism provides a friendly courseware management interface to aid teachers to create new course units, upload courseware to the remote courseware & user portfolio database and delete or modify courseware from the remote courseware & user portfolio database.

The Client Mobile Learning System

The central part shown as Fig. 1 displays the system architecture of the client mobile learning system which includes four intelligent agents and four databases. The learning interface agent aims at providing a flexible learning interface for the legal learners with registered accounts stored in the user account database to interact with the feedback agent, personalized courseware recommendation agent, and personalized vocabulary recommendation agent. It provides a friendly learning interface to show the English news articles recommended from the personalized courseware recommendation agent, a user interface for collecting learner’s learning responses, an English vocabulary learning interface for learning new vocabularies, and a checking vocabulary interface for confirming acquired English vocabularies. The feedback agent aims at collecting learner explicit feedback information that includes the difficulty levels and comprehension degrees of the learned English news articles from the learning interface agent and storing them in the user portfolio database for personalized English news article recommendation. The personalized courseware recommendation agent is in charge of recommending personalized courseware from the local courseware database to individual learner, then evaluates learner reading ability according to learner feedback responses based on the proposed fuzzy Item Response Theory. Finally, to promote the reading ability of English news article by enhancing learner vocabulary ability, the personalized vocabulary recommendation agent extracts the new vocabularies to individual learners from the learned English news articles based on the vocabulary abilities of individual learners measured by the personalized vocabulary learning system (Chen & Chung, 2006) and the difficulty parameters of vocabularies. The personalized vocabulary learning system was presented to promote the learning performances and interests of learners’ English vocabulary learning based on Item Response Theory and learning memory cycle, which can recommend appropriate English vocabularies for learning according to individual learner vocabulary ability and memory cycle. It was successfully implemented on personal digital assistant (PDA) for personalized English vocabulary learning without constraints of time or place by mobile devices.

Basically, if the difficulty parameters of the appeared vocabularies in the learned English news article are larger than the current learner vocabulary ability, then these corresponding English vocabularies will be served as likely new vocabularies to show for the learner. Furthermore, a checking vocabulary interface is used to confirm whether these new vocabularies have been acquired by the learner after the learner finishes the vocabulary learning. In the meanwhile, these acquired vocabularies will be simultaneously recorded in the client and server user portfolio databases through the data synchronized agent to avoid repeated vocabulary learning in the next learning cycle. Currently, the vocabulary database contains over eight thousands vocabularies with corresponding Chinese translations and difficulty parameters collected from the vocabulary repository of General English Proficiency Test in Taiwan. The aim of the GEPT test is to provide a fair and reliable check for each level of ability in English. Currently, the GEPT is divided into five levels with content as appropriate to each level, and each level incorporates listening, reading, writing and speaking components. The elementary, intermediate, and high-intermediate levels are administered twice a year, the advanced level once a year, and the superior level upon request. The following section details the system architecture operating procedure.
The Learning Procedure of the Client Mobile Learning System

Based on the system architecture, the details of system operation procedure of the client mobile learning system are illustrated as Fig. 2, and described as follows:

**Figure 2.** The learning procedure of the client mobile learning system

**Step 1.** Learner logs in the client mobile learning system through the learning interface agent. As a learner logs in the system, the learning interface agent will check his account in the user account database. If the learner has already registered and owned past learning records, the system will get his reading ability from the user portfolio database; otherwise, the learner is viewed as a beginner.

**Step 2.** The personalized courseware recommendation agent gets the contents of English news courseware from the local client courseware database, and then the learning interface agent exhibits the learning contents received from the personalized courseware recommendation agent for the beginner. If the learner has been an experienced learner in our system, the learning interface agent will get his past learning records from the user portfolio database to provide personalized learning services through recommending appropriate English news articles to the learner.

**Step 3.** As the learner clicks a piece of English news title, the learning interface agent will get the content of the clicked English news article with corresponding Chinese translations from the local client courseware database and exhibits it to the learner. To read the content of English news article conveniently, each piece of English news was segmented into many small paragraphs with corresponding Chinese translations for learners. The PIMS system provides two switch learning modes, i.e. English mode and Chinese mode, to learners for English news reading learning.

**Step 4.** The learner needs to reply to two simple questionnaires, i.e. the difficulty levels with three various grading scales (easy, moderate or hard) and the comprehension degree (from zero percent to one hundred percent) for the learned English news article in order to obtain personalized learning services in the next learning cycle. These feedback responses are collected by the feedback agent and stored in the user portfolio database for personalized English news article recommendation.

**Step 5.** The personalized vocabulary recommendation agent extracts the unfamiliar or new vocabularies with corresponding Chinese translations to individual learners according to the learners’ vocabulary abilities measured by the personalized vocabulary learning system (Chen & Chung, 2006).

**Step 6.** The personalized vocabulary recommendation agent will summarize all learned vocabularies using the vocabulary checking interface after the learner finishes the unfamiliar or new vocabularies’ learning. The aim is to perform the confirming process for the acquired vocabularies, then the proposed system will store these acquired vocabularies into the user portfolio database.

**Step 7.** Finally, the personalized courseware recommendation agent will re-evaluate the learner new reading ability according to the explicit feedback responses of two simple questionnaires, then return Step 2 for the next learning cycle.

**English E-News Archive**

The English E-news archive aims at automatically gathering English news articles and extracting metadata contained in news articles from FTV English e-news site utilizing an intelligent crawler agent with metadata extraction mechanism. These retrieval English news metadata including English and Chinese news titles, URL address, date,
and news body was stored in the remote courseware & user portfolio database. Finally, the metadata extraction process automatically extracts the meaningful metadata by the matching technique of regular expression and stores them in the remote courseware & user portfolio database.

Measuring Difficulty of English News Article

Generally, the difficulty of English article can be viewed as readability and the past proposed evaluation schemes relating to readability are over thirty (Flesch, 1948; Klare, 2000). Among the proposed schemes, Flesch’s reading ease formula (Flesch, 1948) presented in 1948 is the most widely used approach to evaluate the difficulty for a piece of English article. Currently, Flesch’s reading ease formula is also employed to evaluate the difficulty of English article in the Microsoft Office Word. However, Flesch’s reading ease formula does not consider the differences of native with non-native English speakers while evaluating the difficulty of English article. That is, the reading difficulties are different for native and non-native English speakers while reading the same piece of English article. Therefore, to evaluate difficulty of English article for non-native speakers more precisely, this study integrates Flesch’s reading ease formula with the proposed difficulty parameter inferred by a fuzzy rule base based on the percentage of vocabulary occurring in various GEPT grading levels to present a novel readability measure for estimating difficulty of English news article. Next, the details of the proposed evaluation scheme are explained in the following subsections.

Measuring the Difficulty of English news by Flesch Reading Ease Formula

Flesch’s reading ease formula employed the average sentence length in a piece of English article and the average number of syllables per word to measure the difficulty of English article. Flesch’s reading ease formula can be formulated as follows:

\[
RE = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)
\]

where the \(RE\) represents the reading ease value, i.e. difficulty, and the range of \(RE\) is between 0 to 100, \(ASL\) is the average sentence length in a piece of English article which can be measured using the total number of vocabularies divided by the total number of sentences, \(ASW\) is the average number of syllables per word which can be measured using the number of syllables divided by the number of vocabularies in a piece of English article.

In Eq. (1), if the measured \(RE\) value is equal to zero, then the English article is viewed as the most difficult for reading. This situation occurs under the measured \(ASL\) and \(ASW\) values are over 37 and 2, respectively. Flesch’s study (1948) indicated that the English article is the most appropriate for reading if the measured \(RE\) value is equal to 65. Besides, to integrate Flesch’s reading ease value with the difficulty parameter of English article inferred by a fuzzy rule base using a linear combination with an adjustable weight, the range of the \(RE\) value must be normalized as the same range with the inferred difficulty (i.e. -3 to +3). The integrated difficulty of English news article will be applied to serve as the key parameter to recommend appropriate English news article for individual learners using the proposed fuzzy Item Response Theory. The normalized formula of the \(RE\) value can be formulated as follows:

\[
RE' = RE \times \left(\frac{6}{100}\right) + 3
\]

where the \(RE'\) represents the normalized \(RE\) value.

The Proposed Scheme for Evaluating Difficulty of English News Article

We can logically infer that most part of learners can easily read the contents of some news article if most part of vocabularies used in this news article belong to the elementary level, even the English news has high \(ASL\) and \(ASW\) values measured by Flesch’s reading ease formula. That is, the difficulty parameter of this English news article should be descended in this situation. However, Flesch’s reading ease formula neglects to consider the difficulties of vocabularies used in an English article while measuring the difficulty of English article. Based on the problem, this section presents a novel scheme to infer the difficulty of English article by a pre-designed fuzzy rule base based on the percentage of vocabulary occurring in various GEPT grading levels.
Computing the Percentages of Vocabulary that Appear in Various GEPT Grading Levels for All Gathered English News

At present, the proposed PIMS system contains about 2792 pieces of English news articles, two thousands one hundred vocabularies of elementary level, two thousands six hundreds vocabularies of intermediate level, and three thousands two hundreds vocabularies of high-intermediate level. To infer the difficulty of an English news article, the proposed PIMS system will compute respectively the percentages of vocabulary occupying in three grading levels for all gathered English news articles.

Determining Fuzzy Membership Functions by the K-means Clustering Algorithm for Fuzzy Rule Base

To infer the difficulties of English news articles by employing the fuzzy inference mechanism based on the percentage of vocabulary occurring in various GEPT grading levels, the used input linguistic variables of the fuzzy inference mechanism are first established herein. In this work, the membership functions used in the fuzzy rule base must be logically determined in advance. The K-means clustering algorithm (Rui & Wunsch, 2005) was applied to determine the centers of the triangle fuzzy membership functions automatically according to the data distribution of the percentage of vocabularies occurring in the gathered English news articles herein. To obtain simple fuzzy rule base for inferring the difficulty of an English news article, this study sets the number of clusters in the K-means clustering algorithm as three. In other words, each considered grading level of occurring vocabulary contains three linguistic terms, i.e. low, moderate, and high, to describe a fuzzy rule. After that, the membership functions of the triangle fuzzy sets were automatically determined according to the cluster centers of three grading levels. Suppose the determined centers of three linguistic terms are assigned notations as 1\(c\), 2\(c\) and 3\(c\) for the vocabulary of elementary level. Figure 3 shows an example for the determined fuzzy membership functions of elementary level.

In addition, the output variable of the fuzzy inference mechanism is the difficulty of an English news article. To explain the designed fuzzy rule base, the simplified representation notations of the input and output linguistic variables are listed in Tables 1 and 2, respectively. Moreover, the defined membership functions for the difficulty of an English news article are shown as Fig. 4.

![Figure 3. An example of the fuzzy membership functions determined by the centers of three linguistic terms for the elementary level](image)

<table>
<thead>
<tr>
<th>Linguistic variable</th>
<th>Representation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowly occurring percentage of vocabulary</td>
<td>Low</td>
</tr>
<tr>
<td>Moderately occurring percentage of vocabulary</td>
<td>Mid</td>
</tr>
<tr>
<td>Highly occurring percentage of vocabulary</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 1. The input linguistic variable for three grading levels of vocabularies

<table>
<thead>
<tr>
<th>Linguistic variable</th>
<th>Representation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Lowly Difficult Degree</td>
<td>VL</td>
</tr>
<tr>
<td>Lowly Difficult Degree</td>
<td>L</td>
</tr>
<tr>
<td>Moderately Difficult Degree</td>
<td>M</td>
</tr>
<tr>
<td>Highly Difficult Degree</td>
<td>H</td>
</tr>
<tr>
<td>Very Highly Difficult Degree</td>
<td>VH</td>
</tr>
</tbody>
</table>

Table 2. The output linguistic variable for the difficulty of an English news article
(3) Designing Fuzzy Rule Base for Inferring the Difficulty of an English News Article

By analyzing the percentage of vocabulary occurring in three various grading levels of GEPT, twenty-seven basic fuzzy rules can be summarized to infer the difficulty of an English news article for the learned English news. Table 3 illustrates the fuzzy rule base designed by English course experts for inferring the difficulty of a piece of English news article.

Table 3. The fuzzy rule base designed by English course experts for inferring difficulty of an English news article

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Antecedent part</th>
<th>Consequent part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>E .Low (\cap) I .Low (\cap) HI .Low</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>2</td>
<td>E .Low (\cap) I .Low (\cap) HI .Mid</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>3</td>
<td>E .Low (\cap) I .Low (\cap) HI .High</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>4</td>
<td>E .Low (\cap) I .Mid (\cap) HI .Low</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>5</td>
<td>E .Low (\cap) I .Mid (\cap) HI .Mid</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>6</td>
<td>E .Low (\cap) I .Mid (\cap) HI .High</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>7</td>
<td>E .Low (\cap) I .High (\cap) HI .Low</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>8</td>
<td>E .Low (\cap) I .High (\cap) HI .Mid</td>
<td>(\Rightarrow) D .H</td>
</tr>
<tr>
<td>9</td>
<td>E .Low (\cap) I .High (\cap) HI .High</td>
<td>(\Rightarrow) D .HL</td>
</tr>
<tr>
<td>10</td>
<td>E .Mid (\cap) I .Low (\cap) HI .Low</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>11</td>
<td>E .Mid (\cap) I .Low (\cap) HI .Mid</td>
<td>(\Rightarrow) D .M</td>
</tr>
<tr>
<td>12</td>
<td>E .Mid (\cap) I .Low (\cap) HI .High</td>
<td>(\Rightarrow) D .H</td>
</tr>
<tr>
<td>13</td>
<td>E .Mid (\cap) I .Mid (\cap) HI .Low</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>14</td>
<td>E .Mid (\cap) I .Mid (\cap) HI .Mid</td>
<td>(\Rightarrow) D .H</td>
</tr>
<tr>
<td>15</td>
<td>E .Mid (\cap) I .Mid (\cap) HI .High</td>
<td>(\Rightarrow) D .HL</td>
</tr>
<tr>
<td>16</td>
<td>E .Mid (\cap) I .High (\cap) HI .Low</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>17</td>
<td>E .Mid (\cap) I .High (\cap) HI .Mid</td>
<td>(\Rightarrow) D .M</td>
</tr>
<tr>
<td>18</td>
<td>E .Mid (\cap) I .High (\cap) HI .High</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>19</td>
<td>E .High (\cap) I .Low (\cap) HI .Low</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>20</td>
<td>E .High (\cap) I .Low (\cap) HI .Mid</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>21</td>
<td>E .High (\cap) I .Low (\cap) HI .High</td>
<td>(\Rightarrow) D .H</td>
</tr>
<tr>
<td>22</td>
<td>E .High (\cap) I .Mid (\cap) HI .Low</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>23</td>
<td>E .High (\cap) I .Mid (\cap) HI .Mid</td>
<td>(\Rightarrow) D .H</td>
</tr>
<tr>
<td>24</td>
<td>E .High (\cap) I .Mid (\cap) HI .High</td>
<td>(\Rightarrow) D .VL</td>
</tr>
<tr>
<td>25</td>
<td>E .High (\cap) I .High (\cap) HI .Low</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>26</td>
<td>E .High (\cap) I .High (\cap) HI .Mid</td>
<td>(\Rightarrow) D .L</td>
</tr>
<tr>
<td>27</td>
<td>E .High (\cap) I .High (\cap) HI .High</td>
<td>(\Rightarrow) D .M</td>
</tr>
</tbody>
</table>

Note: The notations E, I, HI represent respectively the input variables of elementary level vocabulary, intermediate level vocabulary, high-intermediate level vocabulary, and D is the difficulty of an English news article.
(4) Fuzzy Inference for Inferring the Difficulty of an English News Article

This section explains how to infer the difficulty of an English news article according to the designed fuzzy rule base by fuzzy inference. The designed fuzzy production rules are formed by IF-THEN rules represented as follows:

\[
\text{IF } X_1 = A_1 \text{ and } X_2 = A_2 \text{ and } X_3 = A_3 \text{ THEN } Y = B
\]

where \( X_1 \) and \( Y \) denote linguistic variables, and \( A_1 \) and \( B \) represent linguistic terms.

A defuzzification process aims to convert the outcome of fuzzy inference into a crisp value of difficulty. In the fuzzy set theory, the center of gravity (COG) (Lin & George Lee, 1996), which is most widely used defuzzification scheme, calculates the crisp value of difficulty from the most typical values and respective degrees of membership function. Therefore, the defuzzification method of center of gravity is utilized to obtain the crisp value of difficulty of an English news article, and the mathematical formula is shown as follows:

\[
ID = \frac{\sum_{j=1}^{n} \mu_z(z_j)z_j}{\sum_{j=1}^{n} \mu_z(z_j)}
\]

where \( ID \) is the inferred difficulty of an English news article, \( n \) is the number of quantization levels of the output; \( z_j \) denotes the amount of control output at the quantization level \( j \), and \( \mu_z(z_j) \) represents its membership degree in the output fuzzy set \( Z \).

Finally, the final difficulty of an English news article is a linear combination of the normalized RE value measured by Flesch’s reading ease formula and the inferred difficulty of an English news article with an adjustable weight assigned to each parameter, and formulated as follows:

\[
\text{Final Difficulty} = w \times \text{RE'} + (1 - w) \times ID
\]

where the \( \text{RE'} \) represents the normalized RE value, \( ID \) is the inferred difficulty of an English news article, and \( w \) is an adjustable weight.

**Table 4.** The content statistics for a piece of English news article entitled “Chen Considers Control Yuan Head”

<table>
<thead>
<tr>
<th>Chen Considers Control Yuan Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>The total number of sentences</td>
</tr>
<tr>
<td>The total number of vocabularies</td>
</tr>
<tr>
<td>The total number of syllables</td>
</tr>
<tr>
<td>The percentage of elementary level vocabulary</td>
</tr>
<tr>
<td>The percentage of intermediate level vocabulary</td>
</tr>
<tr>
<td>The percentage of high-intermediate level vocabulary</td>
</tr>
<tr>
<td>The percentage of the other vocabularies</td>
</tr>
</tbody>
</table>

An Example for Inferring the Difficulty of a Piece of English News Article

This section gives an example to explain how to infer the difficulty for a piece of English news article in detail. Suppose Table 4 illustrates the content statistics for a piece of English news article entitled “Chen Considers Control Yuan Head”. Next, the detailed procedures for inferring the difficulty of this English news article are described as follows:

Computing the RE value of the English news article by Flesch’s reading ease formula

\[
\text{RE} = 206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})
\]

\[
= 206.835 - (1.015 \times \frac{311}{14}) - (84.6 \times \frac{559}{311})
\]

\[
= 206.835 - 31.1 - 130.59
\]

\[
= 32.262
\]

Normalizing the RE value by Eq. (2)
Computing the difficulty of the English news article by fuzzy inference

The percentages of the used vocabularies for three considered grading levels in this English news article are 0.655, 0.106, and 0.062, respectively. Eight fuzzy rules including rule numbers of 2, 3, 5, 6, 11, 12, 14, and 15 in the fuzzy rule base are triggered for inferring the difficulty. Figure 5 shows the triggered consequent parts of output variable for defuzzification. Here, the defuzzification strategy of center of gravity (COG) (Lin & George Lee, 1996) is employed to convert the outcome of fuzzy inference into a crisp value, and computed as follows:

\[
RE' = RE \times \left( \frac{-6}{100} \right) + 3
\]

\[
= 32.262 \times \left( \frac{-6}{100} \right) + 3
\]

\[
= 1.064
\]

\[
\frac{\sum_{j=1}^{n} \mu_j(z_j)z_j}{\sum_{j=1}^{n} \mu_j(z_j)}
\]

\[
= \frac{0.206 \times (-3) + 0.694 \times (-1.5) + \ldots + 0.305 \times 1.5 + 0 \times 3}{0.206 + 0.694 + \ldots + 0.305 + 0}
\]

\[
= -0.547
\]

Figure 5. The triggered consequent parts of output variable for defuzzification

Determining the final difficulty of the English news article by integrating the normalized RE and the inferred difficulty value under the adjustable weight is set to 0.5

\[
\text{Final Difficulty} = 0.5 \times 1.064 + 0.5 \times (-0.547)
\]

\[
= 0.259
\]

Personalized English News Recommendation

Estimating learner reading ability enables the system to recommend the appropriate English news articles to learners. The personalized courseware recommendation agent estimates learner reading ability using the proposed fuzzy Item Response Theory (FIRT), then calculates the information function values of English news articles to recommend appropriate English news articles to individual learners based on the ranking order of information function values of English news articles. The following subsections first describe how to evaluate learner’s reading ability and recommend appropriate English news articles to individual learners based on the FIRT in detail.

Courseware Modeling and Estimation of Learner’s Reading Ability

To estimate learner’s reading ability, the item characteristic function with a single difficulty parameter proposed by Rasch (Baker, 1992) is first used to model a piece of English news article. The formula of item characteristic function with single difficulty parameter is formulated as follows:
\[ P_j(\theta) = \frac{e^{D(\theta - b_j)}}{1 + e^{D(\theta - b_j)}} \]  

where \( P_j(\theta) \) denotes the probability that learners can completely understand the \( j^{th} \) English news article at a level below their ability level \( \theta \), \( b_j \) is the difficulty of the \( j^{th} \) English news article, and \( D \) is a constant 1.702.

Two methods are widely used in assessing learner’s ability in computerized adaptive testing (CAT) field. They are the maximum likelihood estimation (MLE) and Bayesian estimation approaches (Baker, 1992). Although the procedure of MLE is simple and easily implemented, it has the problem of producing divergent estimations for learner’s ability when the learner gives complete understanding or not understanding responses for all learned courseware during a learning process. The Bayesian estimation procedure always converges for all possible learners’ responses (Baker, 1992). For this reason, the Bayesian estimation procedure is applied to estimate learner’s reading ability in this study. Bock and Mislevy (Baker, 1992) derived the quadrature form to approximately estimate learner’s ability as follows:

\[ \hat{\theta} = \frac{\sum_{k=1}^{g} \theta_k L(u_1, u_2, \ldots, u_n | \theta_k) A(\theta_k)}{\sum_{k=1}^{g} L(u_1, u_2, \ldots, u_n | \theta_k) A(\theta_k)} \]  

where \( \hat{\theta} \) denotes the learner’s reading ability of estimation, \( L(u_1, u_2, \ldots, u_n | \theta_k) \) is the value of likelihood function at a level below their ability level \( \theta_k \) and learner’s responses are \( u_1, u_2, \ldots, u_n \), \( \theta_k \) is the \( k^{th} \) split value of ability in the standard normal distribution, and \( A(\theta_k) \) represents the quadrature weight at a level below their ability level \( \theta_k \).

In Eq. (10), the likelihood function \( L(u_1, u_2, \ldots, u_n | \theta_k) \) can be further described as follows:

\[ L(u_1, u_2, \ldots, u_n | \theta_k) = \prod_{j=1}^{n} P_j(\theta_k)^{u_j} Q_j(\theta_k)^{1-u_j} \]  

where \( P_j(\theta) = \frac{e^{D(\theta - b_j)}}{1 + e^{D(\theta - b_j)}} \), \( Q_j(\theta_k) = 1 - P_j(\theta_k) \), \( P_j(\theta) \) denotes the probability that learners can understand the \( j^{th} \) English news article at a level below their ability level \( \theta_k \), \( Q_j(\theta_k) \) represents the probability that learners cannot understand the \( j^{th} \) English news article at a level below their ability level \( \theta_k \), and \( U_j \) is the completely understanding or not understanding answer obtained from learner feedback to the \( j^{th} \) English news article, i.e. if the answer is completely understanding then \( U_j = 1 \); otherwise, \( U_j = 0 \).

In the system presented here, learner reading abilities are restricted to between –3 and 3. That is, learners with reading ability \( \theta = -3 \) are classified as the poorest, those with ability \( \theta = 0 \) are considered to have moderate abilities, and those with ability \( \theta = 3 \) are regarded as having the best abilities. This system adjusts learners’ reading abilities based on learner feedback responses. However, the Bayesian estimation procedure estimates learner reading ability merely through learner’s crisp responses. It cannot estimate learner reading ability through learner’s fuzzy responses. To solve this limitation, the FIRT derived from the original IRT is presented herein. The proposed FIRT can infer learner reading ability based on the replies to two predefined simple questionnaires, which are collected by the feedback agent. In this approach, the fuzzy inference mechanism (Lin & George Lee, 1996) is applied to infer the understanding degree according to the learner’s feedback responses. Based on the evaluation of understanding degree explained in a later subsection, the learner’s reading ability after learning an English news article can be evaluated as follows:
\[
\theta_{j+1} = \begin{cases} 
\theta_j + (\theta_w - \theta_j) \times u & \text{when } 0 \leq u < 0.45 \\
\theta_j & \text{when } 0.45 \leq u \leq 0.55 \\
\theta_j + (\theta_c - \theta_j) \times u & \text{when } 0.55 < u \leq 1 
\end{cases}
\] (12)

where \( \theta_j \) denotes the estimation value of the learner’s reading ability for the total number of \( j \) accumulatively learned English news articles, \( \theta_{j+1} \) represents the estimation value of the new learner’s reading ability after learning the \((j + 1)\)th English news article, \( \theta_w \) is the estimation value of the learner’s reading ability assuming that the learner cannot completely understand the \((j + 1)\)th English news article, \( \theta_c \) denotes the estimation value of the learner’s reading ability assuming that the learner can completely understand the \((j + 1)\)th English news article, and \( u \) represents the understanding degree inferred by the fuzzy inference mechanism for the \((j + 1)\)th English news article.

Using the estimation method of learner’s reading ability mentioned in Eq. (12), if the learner’s understanding degree inferred by the fuzzy inference mechanism ranges between zero and 0.45, then the result indicates that the learner cannot understand the greater part of the contents of the learned English news article. Therefore, the learner’s reading ability in the learned English news should be descended. That is, he/she needs to learn easier English news articles in order to promote his/her basic knowledge in the learned English news article. Moreover, if the learner understanding degree inferred by the fuzzy inference mechanism is between 0.45 and 0.55, then the result indicates that the learner gives a highly uncertain feedback response. Thus, the learner’s reading ability after reading the learned English news articles has not been tuned. On the other hand, if the learner’s understanding degree inferred by the fuzzy inference mechanism is between 0.55 and 1, then the result denotes that the learner can understand the most part of contents of the learned English news article. Thus, the learner’s reading ability after reading the learned English news articles should be promoted. After evaluating the new learner’s reading ability, the system will evaluate the information function values of all gathered English news articles based on the maximum information approach to find out the most appropriate English news articles for the individual learner.

**Inference of Understanding Degree**

To infer the learner understanding degree by employing the fuzzy inference mechanism according to the learner’s feedback responses, the used input linguistic variables of the fuzzy inference mechanism are established herein. The variables include the feedback agent’s collection of English news article on the difficulty level and the comprehension percentage for the learned English news article. Besides, the output variable of the fuzzy inference mechanism is the understanding degree of the learned English news article. To explain the designed fuzzy knowledge base, the simplified representation notations of the input and output linguistic variables are listed in Tables 5 and 6, respectively. Moreover, the bell-shaped membership function is then applied to describe the linguistic variables of input and output for the fuzzy inference mechanism. The bell-shaped membership function can be formulated as follows:

\[
\mu_{x_i}(x) = e^{-\frac{(x-m_i)^2}{2\sigma_i^2}}
\] (13)

where \( \mu_{x_i}(x) \) denotes the fuzzy degree of the input \( x \) under the linguistic variable \( x_i \), \( m_i \) and \( \sigma_i \) specify the mean and the width of the bell-shaped membership function for the linguistic variable \( x_i \), respectively. The used membership functions of input and output linguistic variables are shown as Figs. 6, 7 and 8.

By analyzing the results of the learner’s feedback responses, nine basic fuzzy rules can be summarized to infer the learner’s understanding degree for the learned English news article. Table 7 illustrates the designed fuzzy rule base for the proposed fuzzy Item Response Theory. To infer the learner’s understanding degree, the reasoning process of Mandani’s minimum fuzzy implication (Lin & George Lee, 1996) is employed to integrate the triggered fuzzy rules. Moreover, the defuzzification method of center of gravity (Lin & George Lee, 1996) is used to obtain the crisp value of learner’s understanding degree in order to evaluate learner’s reading ability herein.
Figure 6. The membership functions for the comprehension percentage of reading

Figure 7. The membership functions of fuzzy singleton for the difficulty level of reading English news article

Figure 8. The membership functions for the learner understanding degree
Table 5. The input linguistic variables for the comprehension percentage of reading and the difficulty level

<table>
<thead>
<tr>
<th>Linguistic variable for the comprehension percentage of reading</th>
<th>Representation method</th>
<th>Linguistic variable for the difficulty level</th>
<th>Representation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowly understanding percentage</td>
<td>LUP</td>
<td>Easy</td>
<td>E</td>
</tr>
<tr>
<td>Moderately understanding percentage</td>
<td>MUP</td>
<td>Moderately</td>
<td>M</td>
</tr>
<tr>
<td>Highly understanding percentage</td>
<td>HUP</td>
<td>Hard</td>
<td>H</td>
</tr>
</tbody>
</table>

Table 6. The output linguistic variable for the learner understanding degree

<table>
<thead>
<tr>
<th>Linguistic variable</th>
<th>Representation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Lowly Understanding Degree</td>
<td>VLUD</td>
</tr>
<tr>
<td>Lowly Understanding Degree</td>
<td>LUD</td>
</tr>
<tr>
<td>Moderately Understanding Degree</td>
<td>MUD</td>
</tr>
<tr>
<td>Highly Understanding Degree</td>
<td>HUD</td>
</tr>
<tr>
<td>Very Highly Understanding Degree</td>
<td>VHUD</td>
</tr>
</tbody>
</table>

Table 7. The designed fuzzy rule base for inferring learner’s understanding degree of the learned English news article

<table>
<thead>
<tr>
<th>Understanding degree of the learned English news article</th>
<th>The understanding percentage of reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>The difficulty level of reading</td>
<td>LUP</td>
</tr>
<tr>
<td>E</td>
<td>MUD</td>
</tr>
<tr>
<td>M</td>
<td>LUD</td>
</tr>
<tr>
<td>H</td>
<td>VLUD</td>
</tr>
</tbody>
</table>

Maximum Information Evaluation and English News Recommendation

In traditional IRT, two approaches are used to recommend appropriate courseware to the learner. They are the maximum information strategy and Bayesian strategy (Baker, 1992), respectively. Maximum information strategy emphasizes that each English news article with the corresponding difficulty parameter exhibits different information to learner’s reading. English news article with higher-information is more suitable to be recommended for the learner. Since the Bayesian strategy is more complicated than the maximum information approach, the maximum information method is applied to recommend appropriate English news article for the proposed FIRT. The maximum information function is defined as follows:

$$I_j(\theta) = \frac{(1.7)^2}{e^{1.7(\theta-b_j)} \left[ 1 + e^{-1.7(\theta-b_j)} \right]^2}$$ (14)

where \(I_j(\theta)\) is the information function value of the \(j^{th}\) English news article at a level below their ability level \(\theta\), \(b_j\) is the difficulty parameter of the \(j^{th}\) English news article.

After calculating the corresponding information values of English news articles, the personalized courseware recommendation agent can recommend a series of English news articles to the learner with reading ability \(\theta\) according to the ranking order of information function values. An English news article with the maximum information function value under learner with reading ability \(\theta\) indicates that the system presented here gives the highest recommendation priority. Whether the learner accepts the recommended English news article with highest recommendation priority or selects the other recommended courseware to do further reading, our system will record learner’s reading paths and learner’s feedback responses into the remote courseware & user portfolio database during learning processes.

Personalized English Vocabulary Recommendation

Since the key factor affecting the reading ability of English news article is often vocabulary ability, how to promote the reading ability of English news article through increasing vocabulary ability is also a key issue herein.
Moreover, learners have different vocabulary abilities and memory cycles for various vocabularies based on their prior knowledge. This PIMS system employed the personalized English vocabulary recommendation agent to correctly identify new or unfamiliar vocabularies of individual learners from the reading English news article in order to enhance vocabulary abilities of individual learners. In this work, another proposed personalized vocabulary learning system (Chen & Chung, 2006) is used to support the personalized English vocabulary recommendation agent in order to measure the vocabulary abilities of individual learners, identify the memory cycles of individual learners for various vocabularies, and provide a personalized vocabulary learning mechanism. In the meanwhile, the personalized vocabulary learning system can also measure the difficulty of vocabulary based on considering the GEPT grading level of vocabulary, the length of vocabulary, and the ratio of the number of characters to the length of phonetic symbols. The range of the difficulty of vocabulary measured by the personalized vocabulary learning system is normalized between -3 to +3, which is the same with the range of the learner’s vocabulary ability. Here, the PIMS system adopts the difficulty parameters of vocabularies from the personalized vocabulary learning system to identify whether the vocabularies appearing in the reading English news article are new or unfamiliar vocabularies to the learner by comparing the difficulties of vocabularies with the vocabulary abilities of individual learners. In the study, if the difficulty parameters of the appeared vocabularies in the learned English news article are larger than the current learner vocabulary ability evaluated by the personalized vocabulary learning system, then the corresponding English vocabularies will be served as likely new vocabularies to the learner. Therefore, the set $R_{ij}$ of the recommended vocabularies for individual learners can be represented as follows:

$$R_{ij} = (C_j \cap A_i) - L_i$$

(15)

where $R_{ij}$ represents the set of the recommended new or unfamiliar vocabularies contained in the $j^{th}$ English news article for the $i^{th}$ learner, $A_i$ is the set of vocabularies that the corresponding difficulty parameters are higher than the $i^{th}$ learner’s vocabulary ability, $C_j$ is the set of all vocabularies contained in the $j^{th}$ English news article, $L_i$ is the set of the acquired vocabularies of the $i^{th}$ learner.

Experiments

This section first introduces the implemented personalized intelligent mobile learning system on PDA for English news reading learning. Moreover, to demonstrate the learning effectiveness of the proposed system for English news reading learning, some university students were invited to participate in the experiment during five weeks. The experimental environment and results are detailed and analyzed as the following subsections.

The System Implementation

Gathered English News Articles and Courseware Management System

The detailed procedures of English news archive have been mentioned in previous section. This study also developed a courseware management system with functionality of English news archive in the server side. Currently, there are totally 2792 news articles stored in the remote courseware & user portfolio database, which can support effectively English news reading learning. Moreover, Fig. 9 shows the difficulty distributions of 2792 archival English news articles evaluated by the proposed difficulty assessment agent of English news. The difficulty parameters of these gathered English news articles are distributed from −1.70 to +2.68. This uniform distribution of difficulty parameters is helpful to recommend appropriate English news article to learners with different English news reading abilities. In Fig. 9, the news article with difficulty level −3 stands for the easiest one, the news article with difficulty level 0 represents the moderate one, and the news article with difficulty level +3 indicates the most difficult one. The results indicate that the gathered English news articles are towards more difficult than general English articles used in daily lives. Therefore, the gathered English news articles are not suitable for reading learning to learners with poor English ability. In our experiments, there are totally 47 English news articles selected as course materials by an experienced English teacher for English news reading learning during five weeks. In order to recommend appropriate English news article to learners with different English news reading abilities, the difficulty parameters of these selected English news articles are averagely distributed from −1.70 to +2.68.
The Designed System Interface

This section details the PIMS system implemented by the developing tool of Microsoft Visual Studio .NET 2003. Figure 10 shows the designed PDA learning interface. First, the learning interface agent gets the titles of English news articles from the local client courseware database and exhibits them to learners if they use legal accounts to login the system. Figure 10(a) displays the list of all English news titles stored in the local client courseware database. After a learner clicks some English news title, the learning interface agent will get the content of the clicked English news article from the local client courseware database and exhibits them to the learner. Figure 10(b) shows the learning content of the clicked English news article and the designed learning feedback interface for learners. In addition, Fig. 10(c) reveals the corresponding Chinese translation of a selected paragraph in a learned English news article. Many scholars argued that providing the translation enhances L2 reading (Cook, 1992; Krashen, 1981), but some researches proposed that reading in an L2 is not a monolingual event and the cognitive influence of one’s L1 on L2 acquisition has been clearly shown (Upton & Lee-Thompson, 2001; Cohen, 1995). Kern’s (1994) study proposed that L2 readers often access to their L1 and use this resource as a strategy to help them comprehend an L2 text. Upton & Lee-Thompson’s study (2001) also indicated that the mental translation defined by Kern (1994) as the “mental reprocessing of L2 words, phrases, or sentences in L1 forms while reading L2 texts” is a variable that influences reading comprehension.

Kern’s study (1994) also found that L2 readers most frequently used mental translation in response to specific obstacles to comprehension, such as unfamiliar words and structures. Additionally, Krashen’s research (1981) concluded that L1 may “substitute” for the acquired L2 as an utterance initiator when the performer has to produce in the target language, but has not acquired enough of the L2 to do this. Therefore, to avoid reading obstacles during learning processes, each piece of English news article was partitioned into several small paragraphs and each paragraph has a corresponding Chinese translation. This function aims at helping learners to assess whether their reading comprehension is correct after they read a selected paragraph in an English news article. In the meanwhile, looking up the corresponding Chinese translation function is optional, thus it can be determined based on learner self-requrement. In addition, the client mobile learning system can extract the unfamiliar or new vocabularies to individual learners according to the learners’ vocabulary abilities measured by the personalized vocabulary learning system (Chen & Chung, 2006) for English vocabulary learning. Figure 10(d) reveals the list of these likely new vocabularies identified by the personalized vocabulary learning system with the corresponding Chinese translations for the learned English news article. Moreover, Fig. 10(e) exhibits the checking interface of vocabularies for confirming the acquired vocabularies of individual learners. Finally, the client mobile learning system will re-evaluate learners’ reading abilities according to the explicit feedback responses of the learned English news articles, then recommend appropriate English news articles to individual learners for the next learning cycle. Figure 10(f)
shows the list of the recommended English news articles ranked by the reading priority order for the next learning cycle.

(a) The list of English news titles
(b) The learning contents of the clicked English news article and the designed feedback interface for learners
(c) The corresponding Chinese translations for the learned English news article
(d) The list of likely new vocabularies identified by the system with the corresponding Chinese translations
(e) The checking interface of the acquired vocabularies
(f) The list of the recommended English news articles ranked by the reading priority order for the next learning cycle

Figure 10. The designed user interface of the personalized intelligent mobile learning system
Research Target and Limitations

In order to assess the learning performance of the proposed personalized intelligent mobile learning system, this study called a class meeting to recruit fifteen third grade university students who were majoring in the Department of English Teaching at National Hualien University of Education to take part in this experiment according to their willingness. In this class meeting, the researchers of the study first gave an oral presentation to explain the research background, motivations, and experimental procedure of the study, and then exhibited the functionalities of the implemented PIMS on PDA device. After that, this study recruited fifteen volunteers from thirty-five students who attended the meeting. Among fifteen participants, there are two male students and thirteen female students, their first language is Chinese, and their ages are between twenty-one to twenty-three years old. Adopting fifteen participants to conduct this experiment is due to the limitation of the number of high price PDA equipment. Therefore, the experimental results mentioned later are not suitable to be broadly inferred or explained to the other cases since this study only adopted small samples. In addition, fifteen participants had no or very minimal experience on using PDA devices before participating in the experiment. This situation could lead to learning obstacle while operating PDA for reading learning. Moreover, all participants were being trained to be an English teacher of elementary school in their university careers. Thus, they had been educated basic teaching skills of English listening, speaking, reading and writing for three years at least. Since all participants were volunteers and were majoring in English teaching department, they were highly interested in using the novel learning tool for English reading learning, thus owning high learning motivations. Additionally, the learning activity of English news reading only provided static articles displayed on PDA for learners’ reading learning without any listening comprehension functions. Therefore, reading learning of English news articles assisted by news announcer pronunciation has not yet been considered in the study.

Experimental Design

The designed experimental procedure for learning performance assessment is displayed as Fig. 11. The testing sheets of pre-test and post-test, which contain respectively ten choice questions of reading test from ten randomly selected English news articles, were designed by an experienced English teacher for the learning performance assessment. The testing sheets of pre-test and post-test were examined by statistical analysis to confirm owning the same difficulty level. The testing time of the reading test is fifty minutes for both the pre-test and post-test.

![Diagram of Experimental Design](image)

Figure 11. The procedure of the experiment

Before performing the experiment, fifteen participants received the training course of two hours to operate PDA and used the proposed PIMS system, and then they were invited to perform a pre-test for assessing their initial English news reading abilities. The circumstance of the PDA operation training and the pre-test in a computer classroom are
exhibited as Figs. 12(a) and 12(b), respectively. After finishing the training course and the pre-test, each learner was distributed a PDA to perform English news reading learning by the proposed PIMS system during five weeks. In other words, the learners can freely use this handheld system at anytime and anywhere for conducting ubiquitous reading learning of English news articles. In the meanwhile, the proposed system can immediately monitor learners’ learning states via the recorded learning portfolios during learning processes. Teachers can also observe the recorded learning portfolios through the user interface of the courseware management system. When a learner never login the system for English reading learning within two days, the proposed system will send an e-mail letter to remind his/her learning states. Five weeks later, these fifteen students were invited to perform a post-test for assessing their English news reading abilities and fill out a pre-designed questionnaire for evaluating their satisfactory degree after learning.

![PDA operation training](image1)
![The pre-test of English news reading ability](image2)

*Figure 12. The training course for PDA operation training and pre-test in a computer classroom*

### Learning Evaluation

In this section, three evaluating procedures including a pre-test, post-test, and questionnaire were performed to assess the learning outcomes for the proposed PIMS system.

#### Learning Performance Evaluation

Figure 13 displays the comparison results of the learning performance for both the pre-test and post-test. The post-test results indicated that 15 participants’ language abilities prior to starting the reading learning activity have obvious difference. To encourage us is that the learning performances of most learners in the post-test are better than the pre-test except the number 05 learner. To analyze the reason, the study found that the number 05 learner had not conducted the post-test with the other fourteen learners together. Hence, she was invited to perform the post-test alone. We observed that she could not concentrate her attention on the post-test and only conduct the post-test about twenty minutes due to unclear mental or personal factors. This may lead to very poor post-test score. Based on the reason, the post-test score of the number 05 learner was served as noise sample, thus was eliminated from our statistical analysis. In addition, Fig. 14 shows the variation curves of reading abilities against learning times for three observing learners with low, moderate, and high reading abilities during the learning process. The experimental results indicate that the promotion speed of reading ability of the learner 02 with low reading ability is slower than the other two learners with moderate, and high reading abilities during the learning process. By contrast, compared with the other two learners with low, and moderate reading abilities, the promotion speed of reading ability of the learner 07 with high reading ability is the fastest one. Importantly, the reading abilities of learners will be descended when learners cannot understand the recommended news articles due to over high difficulty parameter to them until their reading abilities are promoted by the proposed personalized reading learning strategy. However, our experimental results show that three observing learners with different initial reading abilities before learning gradually approach to the same reading ability after performing about the 31th learning times. The result shows that the proposed PIMS can recommend appropriate news articles to individual learners, thus promoting learners’ reading abilities regardless of their initial reading abilities.
Figure 13. The comparisons of learning performance for both the pre-test and post-test scores

Figure 14. The variation curves of reading abilities against learning times for three observing learners with low, moderate, and high reading abilities during the learning process
In order to compare the difference of the English news reading abilities of the learners before and after learning by the proposed PIMS system, SPSS statistical software was used to analyze the results of the pre-test and post-test. Table 8 lists the paired sample statistics information for both the pre-test and post-test (N = 14). As the results of descriptive statistics listed in Table 8, the mean scores of fourteen learners for both the pre-test and post-test are 65 and 76.4, respectively. In the meanwhile, the standard deviation of fourteen learners for both the pre-test and post-test are 16.05 and 13.36, respectively. Table 9 gives the trial results of the paired sample t-test of the pre-test and post-test scores. This study found that the difference of the mean scores between the pre-test and post-test scores is -11.43, the trial results reach the significant level under the degree of freedom is set to 13 (i.e. t = -2.51, p = .026). In other words, after using the proposed PIMS system, the promotion of learners’ learning performances achieves significant level and the mean testing score increases 11.43 points.

Table 8. The paired sample statistics information for both the pre-test and post-test scores (N = 14)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>65</td>
<td>14</td>
<td>16.05</td>
<td>4.29</td>
</tr>
<tr>
<td>Posttest</td>
<td>76.4</td>
<td>14</td>
<td>13.36</td>
<td>3.57</td>
</tr>
</tbody>
</table>

Table 9. The paired sample t-test for both the pre-test and post-test scores

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-11.43</td>
<td>17.03</td>
<td>4.55</td>
<td>-21.26 -1.59</td>
<td>-2.51</td>
<td>13.00</td>
<td>.026</td>
</tr>
</tbody>
</table>

Moreover, in order to understand whether the proposed PIMS system provides different learning performances for learners with various English reading abilities, the K-means clustering scheme (Margaret, 2002) was employed to group all learners into three clusters based on their pre-test scores. Table 10 illustrates the cluster results based on the pre-test scores of learners. Additionally, Table 11 shows the statistical information of three cluster groups. This study found that the low score group has best learning performance than the other two groups regardless of the average amount of reading articles, the average amount of learning vocabularies, or the average progressive score. In other words, the PIMS system provides the most benefit in terms of learning performance promotion for the low score learners in this study.

Table 10. The cluster results based on the pre-test scores of learners

<table>
<thead>
<tr>
<th>Group</th>
<th>Learner ID</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Score Group</td>
<td>Learner 07, Learner 08, Learner 10</td>
<td>83.33</td>
</tr>
<tr>
<td>Moderate Score Group</td>
<td>Learner 01, Learner 03, Learner 04, Learner 05, Learner 09, Learner 11, Learner 13, Learner 14, Learner 15</td>
<td>67.78</td>
</tr>
<tr>
<td>Low Score Group</td>
<td>Learner 02, Learner 06, Learner 12</td>
<td>40.00</td>
</tr>
</tbody>
</table>

Table 11. The statistical information of three cluster groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Low Score Group</th>
<th>Moderate Score Group</th>
<th>High Score Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average reading amount of English news articles</td>
<td>54.7</td>
<td>25.2</td>
<td>39.7</td>
</tr>
<tr>
<td>Average learning amount of vocabularies</td>
<td>57.0</td>
<td>31.9</td>
<td>54.0</td>
</tr>
<tr>
<td>Average progressive score</td>
<td>26.7</td>
<td>4.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Questionnaire Analysis

To evaluate learners’ satisfaction degrees for the proposed PIMS system, referring to Chen’s et al. research (Chen et al., 2007), a questionnaire which involves twenty-one questions distinguished five various question types was designed to measure whether the provided services in the proposed PIMS system satisfy the real requirements of
most learners. The five question types contain the personal information about learners’ learning by PDA, the convenience of the system operation, the investigation of the learners’ learning attitude towards using the proposed learning system, the suggestions from learners’ feedback, and the self-assessment of learners’ English news reading ability before and after using the proposed PIMS system for learning. Table 12 gives a summarization of the descriptions of question types. The evaluation results of satisfaction degree are listed in Table 13. To conveniently observe the evaluating results, the investigation results of “strongly agreed” and “agreed” are merged as “approved”, and the investigation results of “strongly disagreed” and “disagreed” are merged as “disapproved”.

**Table 12. The descriptions of question types**

<table>
<thead>
<tr>
<th>Question Type</th>
<th>The number of questions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Information about Using PDA</td>
<td>4</td>
<td>To get the personal information about learners who participated in the learning activity.</td>
</tr>
<tr>
<td>System Operation</td>
<td>5</td>
<td>Questions related to the user interface and the contents of learning materials.</td>
</tr>
<tr>
<td>Learning Attitude</td>
<td>7</td>
<td>To investigate whether the system can enhance learners’ learning motivations or interests and promote their learning achievements.</td>
</tr>
<tr>
<td>Advantages and Disadvantages of the Proposed System</td>
<td>4</td>
<td>To ask learners for investigating the advantages and disadvantages of the proposed PIMS system.</td>
</tr>
<tr>
<td>Self-assessment</td>
<td>1</td>
<td>To ask learners for self-assessing their English news reading abilities before and after using the proposed PIMS system.</td>
</tr>
</tbody>
</table>

Table 13(a) lists the investigation results of the personal information. The evaluating results listed in Table 13(b) indicate that the satisfaction degree of “approved” achieves 81.32% in terms of system operation. Table 13(c) indicates that 75.21% learners satisfy the promotion of learning attitude and interest. Moreover, Table 13(d) lists the investigation results of advantages and disadvantages of the proposed PIMS system. Most learners agreed that the proposed PIMS system provides a friendly user interface, some helpful mechanisms for aiding English news reading learning as well as promotes learning motivations and interests of English reading learning. Particularly, most learners indicated that reading English news article with a corresponding Chinese translation is helpful to promote comprehension degree while reading an English news article. In the meanwhile, learners also proposed several suggestions to enhance the system functions, such as adding a dictionary for looking up unfamiliar vocabularies, English news report with corresponding pronunciation for assisting the training of listening comprehension, and so on.

**Table 13. The satisfaction evaluation results of questionnaire**

(a) The investigation results of the personal information

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Question</th>
<th>The Number of Learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Information about Using PDA</td>
<td>Do you or your family have PDA or cell phone with PDA?</td>
<td>Yes: 2, No: 13</td>
</tr>
<tr>
<td></td>
<td>Do you use PDA first time?</td>
<td>Yes: 13, No: 2</td>
</tr>
<tr>
<td></td>
<td>Have you ever used PDA for learning?</td>
<td>Yes: 1, No: 14</td>
</tr>
</tbody>
</table>

(b) The investigation results of the system operation

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Question</th>
<th>Satisfaction Degree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Operation</td>
<td>I agree that the PIMS system provides a friendly user interface.</td>
<td>4(26.7%), 9(60.0%), 0(0.0%), 2(13.3%), 0(0.0%)</td>
</tr>
<tr>
<td></td>
<td>I am very clear about the learning procedure of the PIMS system.</td>
<td>5(33.3%), 9(60.0%), 0(0.0%), 1(6.7%), 0(0.0%)</td>
</tr>
</tbody>
</table>
I can completely understand the meaning of learning materials recommended by the PIMS system. 2(13.3%) 11(73.3%) 0(0.0%) 2(13.3%) 0(0.0%)

I think the PIMS system is a beneficial learning tool to assist English learning. 0(0.0%) 10(66.7%) 0(0.0%) 5(33.3%) 0(0.0%)

I agree that learning English by PDA is very convenient because I can perform English learning at any time and place. 2(13.3%) 9(60.0%) 0(0.0%) 4(26.7%) 0(0.0%)

Average  81.32% 0% 18.66%

(c) The investigation results of the learning attitude

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Question</th>
<th>Satisfaction Degree (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>strongly agreed</td>
</tr>
<tr>
<td>Learning Attitude</td>
<td>I agree that the design learning materials on the PIMS system can promote my learning interests.</td>
<td>0(0.0%)</td>
</tr>
<tr>
<td></td>
<td>I often increase my learning time because learning by the proposed PIMS system promotes my learning interests.</td>
<td>2(13.3%)</td>
</tr>
<tr>
<td></td>
<td>I think that using the PIMS system can effectively promote my English news reading ability.</td>
<td>3(20.0%)</td>
</tr>
<tr>
<td></td>
<td>The self-inspection interface provided by the PIMS system can encourage my learning motivation.</td>
<td>3(20.0%)</td>
</tr>
<tr>
<td></td>
<td>I agree that using PDA to learn English is a very interesting learning mode.</td>
<td>2(13.3%)</td>
</tr>
<tr>
<td></td>
<td>I agree that the PIMS system can recommend English news articles with appropriate difficult to me.</td>
<td>2(13.3%)</td>
</tr>
<tr>
<td></td>
<td>I think that the English news articles with the corresponding Chinese translations are helpful to English news reading learning for Chinese students.</td>
<td>4(26.7%)</td>
</tr>
</tbody>
</table>

Average  75.21% 3.82% 20.98%

(d) The summarized investigation results from the learners’ feedback responses

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
</table>

1. The PIMS system provides a friendly learner interface.
2. The PIMS system is helpful to me to know immediate news events except English news reading learning.
3. The PIMS system is conveniently carried for any time and place learning.
4. The PIMS system can promote learning motivations and interests of English learning.
5. Reading English news article with a corresponding Chinese translation is helpful to promote comprehension degree for the reading news article.
6. The self-inspection function of learning processes is helpful to me to understand my learning states and inspire my learning motivation.
7. The adopted English course materials should be more broadly. For example, including English funny stories or English delight articles.
8. English news articles should add the function of news report to assist the training of listening comprehension.
9. After finishing reading learning, the PIMS system can provide the reading test function except self-assessment feedback. This will be more helpful to evaluate reading comprehension degree.
10. The PIMS system should add dictionary function for supporting vocabulary learning.
11. The English news articles that have been read should be remarked or filtered out to avoid repeat learning.
12. The wireless function for connecting with the courseware server is unstable sometime.
Finally, this study also designed a self-assessment question for self-evaluating learners’ English news reading abilities of before and after learning. The question asked learners to assess their English news reading abilities before and after learning based on a score with ten different scales. This study also used SPSS statistical software to perform the paired sample t-test in order to evaluate the difference of self-assessing learning performance. Tables 14 and 15 display the paired sample statistics and the paired sample t-test results of the self-assessing English news reading abilities of learners before and after learning, respectively. The paired sample t-test results show that the difference of learning performance reaches significant level (i.e. $t = -6.25$, $p = .000$). That is, most learners thought that their reading abilities of English news are obviously promoted after using the PIMS system for English news reading learning.

Table 14. The paired sample statistics of the self-assessing English news reading abilities of learners before and after learning ($N = 15$)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>6.67</td>
<td>15.00</td>
<td>1.14</td>
<td>0.30</td>
</tr>
<tr>
<td>After</td>
<td>7.70</td>
<td>15.00</td>
<td>0.92</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 15. The paired sample t-test results of the self-assessing English news reading abilities of learners before and after learning

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before - After</td>
<td>-1.03</td>
<td>0.64</td>
<td>0.17</td>
<td>-1.39</td>
<td>-0.68</td>
<td>-6.25</td>
<td>.000</td>
</tr>
</tbody>
</table>

Research Discussion

In this section, how to apply the proposed scheme to the reading learning of any languages is first discussed. Additionally, the characteristics of the proposed mobile-based personalized reading learning scheme are critically examined and compared with several past research findings of computer-assisted reading learning. Moreover, the ownership issue is addressed to emphasize that it is one of five critical success features while performing mobile learning. Finally, mobile learning associated with learning context is also discussed herein.

How to Apply the Proposed Scheme to the Reading Learning of any Second Languages

In this study, the proposed PIMS system can recommend appropriate English news articles to individual learners according to their reading abilities estimated by the proposed fuzzy Item Response Theory. The experimental results demonstrated that the personalized reading learning of English news articles with enhancing unfamiliar vocabularies facilitates learners' reading abilities. Actually, the recommended learning article by the proposed PIMS system could be any kinds of language article with a corresponding difficulty parameter. Certainly, how to precisely evaluate the difficulty of a reading article to any considered second languages is the most critical issue for the proposed scheme of aiding reading learning. Thus, the proposed learning mechanism can be adopted as an effective strategy for the reading learning of any second languages except for English reading learning. Moreover, the proposed PIMS system implemented on mobile devices is beneficial to reading learning anytime, anywhere, thus enabling learners utilizing their spare time for reading learning.

Comparison of the Research Results with the Past Researches in L2 Reading and Vocabulary Acquisition

In this study, the experimental results indicated that English news reading learning accompanied with unfamiliar or new vocabulary learning provides benefit to promote reading comprehension and learners’ reading abilities. The results confirm those of Huckin et al. (1993) and Alderson (1984), who claimed that vocabulary and vocabulary knowledge is helpful to promote reading comprehension as well as vocabulary is the main factor affecting reading performance in the second language. Moreover, up to 86.7% of learners who participated in the experiment agreed that English news reading learning with assistance of the corresponding Chinese translation provides benefits in terms of English news reading learning for Chinese students. Unlike the subject on “first-language interference” claimed by some scholars (Yorio, 1971; Cook, 1992; Krashen, 1981), this result of the study tends to support that accessing L1 is an assisted learning strategy to help L2 learners comprehend an L2 text. Furthermore, the promotion
of learners’ reading learning performances in the study also confirms that the proposed personalized reading learning strategy based on individual reading abilities indeed helps learners to speed up reading performance and to reduce reading cognitive overload during the learning process. The result verifies that correct reading strategies emphasized by Alderson (1984) enable learners’ successes in reading foreign language. Finally, the proposed PIMS aids learners to conduct ubiquitous language learning at any time from any place by mobile devices. This mechanism is very convenient to utilize learner’s spare time for language learning because common advice from many English teachers on the best way for learning English is “little and often” (Buda, 1984).

Ownership of Mobile Learning

Naismith and Corlett (2006) surveyed many successful mobile learning projects in the proceedings of the mLearn conferences from 2002-2005, and identified five critical success features. One of five crucial factors mentioned in the study is ownership. From the point of view, learners will become more motivational, more active in communication and learn much better when they either own the novel learning tool or treat it as if they own it (Luckin et al., 2004; Attewell & Webster, 2005; Facer et al., 2005). In this study, each learner was distributed a PDA for English reading learning during five weeks. Before performing the experiment, all learners were told that they could use PDA whatever they wanted beyond English learning purposes. It means that learners can install any other software, be free to customize their PDAs, or even use it subversively, not just using PIMS system for English learning. According to our observations from the experiment, we found that most learners just followed our instructions and guidelines to use PIMS system everyday, not using the other features of PDA. However, if these 15 learners who treated PDA more as their own stuffs, then they would get great achievements in the activity of mobile English reading learning. Except the ownership of mobile equipment, the ownership of English learning is also a critical issue for learners to learn English well. In the study, all participants were volunteers and were asked to use PISM system in their school lives, not related to any English courses that they were majoring in. Nearly at the end of experiment, the learning activation in mobile English learning activity was gradually descending, because the date of school exam was approaching and they were being busy in preparing their midterm exams. We found that learners viewed mobile English learning as an extra work, not took it as their own English learning. Thus, how to support learners to use assisted learning tools and enhance their learning ownership in mobile learning, especially in English learning, is an essential issue. Although the results of questionnaire show that learners felt the PIMS system can promote their learning motivations and interests, our future efforts should focus on how to make learners view learning English as their own.

Pedagogy and Context of English Learning

In this study, the experiment was not conducted in a classroom environment, and all learners learnt English by the PIMS system as informal learning in their daily lives. Novel assisted tools that enable learners to perform new activities may change the way learners perceive and practice old activities, and may give rise to additional, unpredicted patterns of learning (Sharples et al., 2007). Therefore, which mobile learning pedagogy is more benefit for English reading learning, what role a teacher should play for and how the interaction should be conducted in a classroom learning need to be further investigated. Additionally, the context of English learning is also important. In our experiment, one of 15 participators indicated that she preferred to use the PIMS system and learnt English in the school library or in the school restaurant, but the reading materials provided by the system were too serious, and not fit with her learning context. She hoped to read articles more related to her learning environment. However, the PIMS system provides great mobility for learners to learn English, but more efforts should consider how to appropriately recommend English reading materials to learners according to learners’ learning contexts and support learners to construct knowledge from learning context.

Conclusion

This study proposes a personalized intelligent mobile learning system (PIMS) to promote the reading ability of English news for individual learners based on the proposed fuzzy Item Response Theory, which can estimate the reading abilities of individual learners and recommend appropriate English news articles to individual learners according to learners’ simple feedback responses and difficulty parameters of the learned English news articles. Compared to the original Item Response Theory, the FIRT is superior to IRT because it can process non-crisp
response to correctly estimate learner’s ability via the revised estimating function of learner’s ability. Moreover, difficulty parameters of English news articles can also be appropriately determined by the proposed measuring scheme of difficulty parameter for non-native English speakers who treat English as second language. To enhance vocabulary learning while reading English news articles, the unknown or unfamiliar vocabularies of individual learners can also be automatically discovered and retrieved from the reading English news articles by the PIMS system according to the English vocabulary abilities of individual learners. Experimental results confirmed that recommending appropriate learning articles to individual learners provide benefits in terms of reducing learner cognitive overload during learning processes, thus promoting learning effects and interests. Meanwhile, the experimental results also indicated that English news reading learning accompanied with unfamiliar or new vocabulary learning provides benefit to promote reading comprehension and learners’ reading abilities. In addition, the study tends to support that accessing L1 is an assisted learning strategy to help L2 learners comprehend an L2 text. More significantly, the proposed PIMS system enables a seamless ubiquitous learning environment for English learning at any time from any place by mobile devices.

Although the proposed PIMS provides benefits in terms of the promotion of English reading learning, there are several issues which are valuable to be further investigated. First of all, except news contents, providing diversified learning contents related to daily-life issues could further promote learning interests of learners while using the PIMS for reading learning. Secondly, adding the reading function accompanying with news announcer pronunciation may be helpful to simultaneously promote learner reading performance as well as listening comprehension. Additionally, the self-assessing feedback responses of reading comprehension of the learned news article replaced by reading test is another considered issue because it can effectively promote the accuracy of evaluating learners’ reading abilities. Moreover, adding dictionary function for supporting vocabulary learning while reading learning has been considered as our future work according to the feedback responses of participants. More importantly, how to enhance learners’ ownership while using PIMS for English reading learning should be viewed as an essential issue. Finally, developing a personalized context-aware ubiquitous English reading learning system, which can recommend appropriate reading learning contents associated with context-awareness information of individual learners, is also our future research issue.

Acknowledgement

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 96-2520-S-004-001.

References


Effects of the Asynchronous Web-Based Course: Preservice Teachers’ Achievement, Metacognition, and Attitudes towards the Course

Abdullah Topcu
Measurement and Evaluation Office Army School of Language, Turkey // Tel: +90 216 518 96 81 // abdullah.topcu@boun.edu.tr

Behiye Ubuz
Faculty of Education, Middle East Technical University, Turkey // Tel: +90 312 210 40 62 // ubuz@metu.edu.tr

ABSTRACT
This present study aimed to investigate the effects of the asynchronous web-based “General Teaching Methods” course conducted based on direct or indirect instructional methods on pre-service teachers’ achievement, metacognition, and attitudes towards web-based course. Two groups, a total of 63 third grade elementary science and mathematics education pre-service teachers, attended to one of the aforementioned web-based courses. The results indicated that the group exposed to the web-based indirect instruction significantly outperformed the group exposed to the web-based direct instruction in the final exam but not in the midterm exam and in the metacognitive questionnaire. On the other hand, the group exposed to the web-based direct instruction had significantly higher attitudes towards web-based course than the one exposed to the web-based indirect instruction.

Keywords
Direct instruction, Indirect instruction, Web-based course, Achievement, Attitude, Metacognition

Introduction
Considerable evidence exists to support the claim that web-based direct instruction (WBDI) or web-based indirect instruction (WBII) compared to the face-to-face (F2F) learning environment promotes students’ achievement (e.g. Matuga, 2001; Pevato, 2003), attitudes toward courses (e.g. Hislop, 2000; Richardson & Price, 2003), and metacognition (e.g. McLoughlin & Luca, 2002; Norton, 2005). The important inflection point for a web-based instruction (WBI), however, is determining which methods are best for a given subject matter (Merisotis & Phipps, 1999). Therefore, there has been an upsurge in publications (Gagne & Shepherd, 2001; Russell, 1999) on the need for instructional theories and their validation in web-based environment as in the F2F learning environment. Having established these facts, the aim of this study is to investigate the effects of the instructional methods (direct or indirect) in the asynchronous web-based course on the preservice teachers’ achievement, metacognition and attitudes towards web-based course (WBC). Specifically, the study sought to address the following research questions: (1) What is the effect of the asynchronous WBII on the pre-service teachers’ achievement in the “General Teaching Methods” course?; (2) What is the effect of the asynchronous WBDI and WBII on the pre-service teachers’ attitudes towards WBC?; and (3) What is the effect of the asynchronous WBDI in pre-service teachers’ metacognition?

Jonassen (2000) proposed that there were four primary instructional goals of WBDI. These are presenting course content in a manner that it hierarchically structures the sequence of information, obtaining student feedback to ensure accuracy of understanding, providing opportunities for students to question the instructor in order to ensure accuracy of understanding and, lastly, creating opportunities for students to communicate with each other in order to share their understanding of course content.

In contrast to the WBDI, Jonassen (2000) proposed that the primary instructional goals of WBII are: Presenting a problem-solving situation in a realistic context, providing opportunities for learners to collaboratively construct knowledge based on multiple perspectives, discussion, and reflection, providing opportunities for learners to articulate and revise their thinking in order to ensure the accuracy of knowledge construction, and, lastly, creating opportunities for the instructor to coach and facilitate construction of student knowledge.

ISSN 1436-4522 (online) and 1176-3647 (print). © International Forum of Educational Technology & Society (IFETS). The authors and the forum jointly retain the copyright of the articles. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the full citation on the first page. Copyrights for components of this work owned by others than IFETS must be honoured. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from the editors at kinshuk@ieee.org.
The Effects of Asynchronous WBI on Achievement

Student achievement in a learning environment is an important learning outcome often examined by researchers. The effect of WBDI (Keller, 1997; Manathunga, 2002; Manuel, 2001; Miller, 2000; Pevato, 2003; Ryan, 2000) or WBII (Gagné & Shepherd, 2001; Jonassen, 2000; Matuga, 2001; Sener & Stover, 2000; Serban, 2000; Tam, 2000; Wegner, Holloway, & Gordon, 1999; Willis, 1995) compared to F2F on educational major students achievement has been investigated widely by researchers.

Studies conducted on comparing WBDI with F2F revealed that WBDI groups performed significantly higher than did F2F groups (Pevato, 2003). The reasons for this significant effect were reported as the highly structured and easily reached content of the course. In other studies, no significant difference was found on students’ achievement exposed to either WBDI or F2F environment (Keller, 1997; Manathunga, 2002; Miller, 2000; Ryan, 2000). These environments did not provide for students a perspective to outperform with respect to other as objectives guide development in both environments. Despite that linear structure of the WBDI course provided students with the feeling of control for their learning process and it caused high motivation to study harder though the students were separated from their teachers by space and time. Manuel (2001), however, reported that F2F group students outperformed better than WBDI group students, as WBDI group students did not significantly master the content due to a lack of experience in the web technologies and the non-interactive structure of the course.

WBII has been proposed as an effective instructional method by many theorists and researchers and assumed as the main stream in the web-based learning environment (Jonassen, 2000; Tam, 2000; Willis, 1995). Matuga (2001) reported in his study that students exposed to the WBII were significantly better than students exposed to the F2F instruction, as WBII encouraged students to think critically and provided students more interaction with their friends so that they felt the community of learning. Other studies (Gagné & Shepherd, 2001; Sener & Stover, 2000; Serban, 2000; Wegner et al, 1999) led to a conclusion of no significant difference between WBII and F2F instruction on student achievement. The core reason for this finding was that students needed strong metacognitive abilities in the WBII.

Although there are no studies that have focused on the effectiveness of WBII compared to WBDI in higher education, the bulk of the evidence in the studies above led us to predict that students using the WBII would score significantly higher on the achievement measures than those who study the course in WBDI.

The Effects of Asynchronous WBI on the Attitude towards WBC

Another important outcome often examined by researchers is students’ attitudes towards WBC (Fredericksen, Pickett, Pelz, Shea, & Swan, 2000; Hislop, 2000; Howland & Moore, 2002; Larson & Bruning, 1996; Lawless, 2000; Oliver & Omari, 2001; Richardson & Price, 2003; Taylor & Mohr, 2001; Weems, 2002). Generally speaking, students’ attitudes towards WBC are affected by their previous computer experiences, their familiarity with the web-based tools (Internet, email, forum, Listserv, etc.), their attitudes towards the course itself, instructional methods used in the course, the content of the course, and a lack of the physical existence of their friends.

Studies on educational major students’ attitudes towards WBC in WBDI environment reported various findings. Hislop (2000), Richardson and Price (2003), and Taylor and Mohr (2001) found that students’ attitude towards WBC increased as most of them expressed high level of satisfaction with the course, and improved confidence in present and future courses. In contrast, Larson and Bruning (1996), and Weems (2002) found that the students’ attitude towards WBC did not change significantly. The no significant change in these findings is due to the fact that the students could not interact with their friends sufficiently to build a learning community. Therefore they found the learning environment static and felt themselves less confident about the accuracy of their knowledge.

Studies on educational major students’ attitudes towards WBC in WBII environment reported that students might have or not have a positive attitude towards WBC, but their attitudes did not significantly change or develop throughout the course period. Howland and Moore (2002), and Fredericksen et al (2000) reported that students had positive attitudes about their online course experience. But, their attitude toward WBC did not change significantly through the course period. Although some students described their learning experience as rewarding but challenging,
others were overwhelmed with the need of relying on themselves. Similarly, the studies of Oliver and Omari (2001) and Lawless (2000) reported that students did not develop a significant positive attitude toward WBC because the course requirements were very challenging and time consuming.

The bulk of evidence in the studies above led us to predict that participants using the WBDI would score significantly higher on the attitude measure than those studying the course in WBII environment.

The Effects of the Asynchronous WBI on the Metacognition

Metacognition generally refers to the ability to understand, control, monitor, and manipulate individual cognitive processes (Brown, Bransford, Ferrara, & Campione, 1983; Reeve & Brown, 1984). The relations of the metacognition with autonomy and interaction push the attention of the researchers on the developments of the students’ metacognitive skills in the web-based learning environment (Tsai, 2004). There might be cyclic relations between some elements of the web-based learning environment and students’ metacognition (Moore, 2002). In other words, web-based learning environment tools and elements stimulate the use of the metacognitive skills and metacognitive skills depending on its levels facilitate interactions and meaningful learning in this environment.

Students can construct knowledge from various and vast store of information by the help of metacognitive knowledge, monitoring and control in the web-based learning environment (Schraw, 2000; Tsai, 2004). Besides, McLoughlin & Luca (2002) reported that web-based learning is likely to be more metacognitive and self-directed. For example, the question prompts and forum discussions in the web-based environment have been found effective to help students focus attention and monitor their learning through elaboration on the questions asked (Norton, 2005). So Asynchronous WBC tools and elements (help functions (Renkl, 2002); pop-up windows (Thomas, 2002); forum discussions (Garrison & Anderson, 2003); preview windows (Cress & Knabel, 2003) might have effects on the development or improvement of the students’ metacognition whether or not instructional methods are direct or indirect.

Methods

Sample

The accessible population of the study was all third grade mathematics and science pre-service teachers in the department of elementary education at public universities in Turkey in which the medium of instruction is English. There are two such universities: Bogazici and Middle East Technical University, having altogether 133 third grade mathematics and science pre-service teachers.

Each university was invited to participate in the present study. Only Bogazici University responded positively to the invitation. Sixty-three pre-service teachers taking “General Teaching Methods” course in their third year of the educational program constituted the sample of this study.

The pre-service teachers were distributed identically to the two groups (WBDI group and WBII group) considering their grade point average (GPA), gender, age, and program (mathematics or science). The members of each matched pair were then assigned to the groups at random. The pre-service teachers distributed to the WBDI group (WBDIG) and WBII group (WBIIG) were 31 and 32, respectively. The 32 WBIIG were composed of 12 male and 20 female, whereas the 31 WBDIG were composed of 13 male and 18 female. The number of pre-service teachers from mathematics was 17 for WBIIG and 18 for WBDIG. The ages of the sample ranged from 21 to 25. The mean of the ages was about 22 for both groups. Of the pre-service teachers participated in the study, their GPA ranged from 1.63 to 3.74 with a mean of 2.41. The mean of the GPA of the WBDIG was 2.40 whereas that of the WBIIG was 2.41.

In analyzing pre-service teachers’ prior experience in computer, 92 % of the pre-service teachers had previously completed at least one computer related course, such as Introduction to Computer, or Computer-Assisted Instruction. Moreover, about three fourths of the pre-service teachers from both groups described themselves as having sufficient keyboarding skills.
Measuring Tools

Midterm and Final Exams of Teaching Methods Course

The midterm exam (ME) and the final exam (FE) required the participants to answer the questions related to the learning objectives addressed in the first five chapters and in the all ten chapters of the one-semester course, respectively. They served to measure the pre-service teachers’ achievement in the “General Teaching Method” course. The 32 questions for the ME and the 34 questions for the FE were submitted to a three-member validation panel, which was composed of the lecturer (the second researcher), who had taught this course for eight years and two English teachers, who had taken similar course previously and had ten years of teaching experiences. The lecturer assessed the questions regarding the appropriateness to the content of the related chapters in the course, the clarity, the correct usage of the language, and the cognitive levels measured whereas English teachers assessed the questions only for the correct usage of the language, their readability and clarity. Most of these questions were directly taken or were slightly adapted from different sources: item pools of the validation panel lecturer, teaching method books (Borich, 1997; Chambers & Sprecher, 1996; Clark & Starr, 1996; Cruickshank, Bainer, & Metcalf, 1995), and the Internet sources (some of the 24 sites are Cotton (2001, sections 1.3), Jarrett (2000). The recommendations of the panel were used to modify and to select the questions for the ME and the FE. Sixteen questions were chosen for the ME, whereas 21 questions were chosen for the FE. Twenty two percent of the FE questions were chosen from the first five chapters. In other words, they were from the content of the ME.

The questions in the ME and the FE were in the form of True/False (Three questions in the ME and four questions in the FE), Fill-in-the-Blanks (two questions in the ME and three questions in the FE), Matching (two questions in the ME and two questions in the FE), Multiple-Choice (five questions in the ME and five questions in the FE), and open-ended (four questions in the ME and seven questions in the FE). The variety of question types in ME or FE aims to measure achievement in the course properly, because each question type requires different cognitive levels and learning styles. By this way, researchers wanted not to leave any pre-service teachers outside the measurement process. Besides, they wanted to increase the validity of the instruments. Specimen questions for each type from the ME and the FE are given in Appendix.

The questions considered in the ME endeavour to test knowledge and understanding of questioning skills, question types, key behaviors in effective teaching, multiple intelligence, problem solving stages, direct instruction, expository teaching, and discovery teaching. The questions considered in the FE endeavour to test knowledge and understanding of discussions, cooperative learning, project-based instruction, computer-assisted instruction, and drama in addition to the topics covered in ME.

The first researcher scored all the answers except the ones given to the open-ended questions. True/false, Matching, Fill-in-the-blanks and multiple-choice items were scored out of two. On the other hand, two researchers scored answers for the open-ended questions separately and they looked for a consensus at their scoring with respect to pre-prepared answer key. They were scored out of six, eight or more with respect to their requirements of higher level thought, critical thinking and effort. For instance, the open-ended question in FE given in Appendix was eight points. Writing “dramatic enactment” truly was one point, finding “dramatic moments” truly was three points, writing “communication forms” truly was two points and writing “representation forms” was two points. The total score of the ME and the FE was 100. The Cronbach alpha values for the internal consistency reliability for the ME and the FE were calculated as .71 and .78, respectively.

Web-Based Course Attitude Scale

The Web-Based Course Attitude Scale (WBAS) (Ham, 2002) updated from the originally developed scale (Hiltz, 1994) was used to determine pre-service teachers’ attitude towards web-based course. WBAS is a five-point Likert-type scale consisted of 47 items measuring (1) Current Feelings about WWW (e.g. stimulating – dull; Fun-Dreary); (2) Online Course Support (e.g. I needed a lot of help to access course materials on the web); (3) Level of Communication (e.g. Online collaborative activities took too much of my time;) and (4) Perception of Satisfaction and Success (e.g. Taking a web-based course is more boring; I spent too much time surfing on the web instead of studying). The sum of the choices for each item yields a raw score ranging from 47 to 235. In addition, it includes
eight survey-type questions to obtain pre-service teachers’ feelings about the online course and their motivation level for the success.

Internal consistency reliability (alpha) estimates for the WBAS itself and its subscales calculated on the basis of the post-test scores were as follows:

- Current Feelings about WWW: .82
- Online Course Support: .70
- Level of Communication: .83
- Perception of Satisfaction and Success: .71
- WBAS total: .90

General Metacognition Questionnaire

The General Metacognition Questionnaire (GMQ) covering all the components of metacognition -metacognitive knowledge, metacognitive judgments and monitoring, and self-regulation, and control of cognition- without tightening them to any specific domain, such as mathematics, science, literature - was used to determine the level of the metacognitive abilities of the participants. GMQ is a five-point Likert scale developed by the authors. Students were asked to rate 30 statements on the perspective of the judgments about previous learning condition. The authors in a study of 221 pre-service teachers investigated the psychometric properties of the GMQ and found that it was sufficient for use in this study. An overall review of validity data (exploratory analysis extracts factor model and this model emerged from the exploratory analysis was confirmed by confirmatory factor analysis) suggests that the GMQ measures what it aims to measure. GMQ scores could range from 30 to 150. Higher scores indicate higher metacognition ability. Internal consistency reliability estimate of .79 was obtained for this questionnaire on the basis of the posttest scores. Examples of items related to each component –metacognitive knowledge, metacognitive judgment and monitoring, self-regulation, and control- include respectively: (a) When I study, I practice saying to myself the important facts, principles and concepts in my cognition, (b) When a different idea comes into my mind instantaneously, while I am studying, I check it in a different condition, and (c) If I do not comprehend one point in a subject, I return to the related section or words.

Observation Checklist

To determine whether web-based asynchronous “General Teaching Methods” course conducted according to the described instructional methods or not, an observation checklist (OC) was developed by the authors. The items include the most important characteristics of the WBII and WBDI (Jonassen, 2000; Tam, 2000; Willis, 1995). Each of the items uses a five-point scale. Items contain instructor (four items), students (seven items) and medium (four items) related criteria. Examples of the items related to each criterion include (a) “Instructor acts as coach, not a content provider”, (b) “Students focus on knowledge building rather than knowledge reproduction”, (c) “Course site provides links related to the course content to stimulate recall for the prior learning”, respectively.

Data Collection and Analysis

Prior to the beginning of the treatment, GMQ and the first subscale, “Current Feelings about WWW” (CFW), of WBAS were posted to the pre-service teachers by email attachment to be returned in one-week time. The ME and the FE were administered in two different classrooms at the middle and at the end of the semester, respectively. One research assistant was present in each classroom. The instructor (the first researcher) acted as controller in the classrooms. WBDIG and WBIIG pre-service teachers were distributed randomly into the two classrooms considering that each classroom had equal number of pre-service teachers from the two groups. The WBAS and the GMQ were also administered by email attachment at the end of the semester.

The dependent variables of this particular design are pre-service teachers’ “Teaching Method Course” performance in ME, and in FE, their WBAS, their post current feelings about WWW (POSTCFW), and their post metacognition (POSTGMQ). These dependent variables were determined as continuous variables and measured on interval scale. The independent variables of the study were collected in two groups as covariates and group membership (main
effect); Block 1 and Block 2. Pre-service teachers’ age, gender, program, GPA and their pre feelings about WWW (PRECFW), their pre metacognitive ability level (PREGMQ) within Block 1 as covariates. Treatment (Methods of instruction) was included in Block 2 as main effect. The treatment and pre-service teachers’ gender and program were discrete variables and were measured on the nominal scale, whereas their GPA, PRECFW and PREGMQ were continuous variables and were measured on interval scale. The pre-service teachers’ gender was coded as zero for female and one for male. The program was coded as one for math education and two for science education. Lastly, treatment was coded as one for WBDI and two for WBII.

Statistical technique named multivariate analysis of covariance (MANCOVA) was used since it is an extension of analysis of covariance that incorporates two or more dependent variables in the same variables (Fraenkel & Wallen, 1996). The design of the study was a single-factor design. The treatment, independent variable; WBII and WBDI, had to levels. Variance due to the GPA, PRECFW, PREGMQ, gender, program, and age was removed prior to entry of the treatment variable. Block 1 (covariates) was entered first in the MANCOVA model. Block 2 (group membership-treatment) was entered second in the analysis while Block 1X2 (covariate*group interaction) was entered third to determine covariate*group membership interactions. This set must be statistically non-significant for MANCOVA model to be valid. Block 1X2 yielded non-significant increase in total variance for the overall MANCOVA model. Therefore, the interaction set was discarded from the inferential statistical analysis. After MANCOVA analysis, follow-up ANCOVA’s (Protected Univariate F Tests) were used for significant main effects. Step-down analyses were conducted for the significant main effects to remove the effects of the relation between them.

Treatment

“General Teaching Methods” web-based asynchronous course includes ten chapters. These are, in sequence, (1) Introduction to Course and The Effective Teacher, (2) Questioning in the Classroom, (3) Introduction to Methods and Direct Instruction, (4) Discovery Teaching, (5) Problem Solving, (6) Discussion, (7) Cooperative Learning, (8) Computer Assisted Instruction, (9) Project-Based Learning, and (10) Drama. Video clips were used for chapters 1, 5, 6, 7 and 8, and the episodes in terms of dialogs were used for the rest of the chapters as an ill-structured case. Each chapter lasted one week except the first one. The first chapter continued one and half weeks as the pre-service teachers got accustomed to the web-based asynchronous course environment and its tools.

Figure 1 and Figure 2 show the flow of the courses for each week in the WBDIG and WBIIG, respectively. The difference between WBDI and WBII was sourced from timing in the presentation of the lecture notes and the questions directed in the discussion forum. The numbers of periods, threads, and course topics were the same for both groups. In these groups, three or four threads were opened in the first two periods and under each thread three or four questions were directed to the pre-service teachers.

In WBDIG, lecture note related to each week was presented at the beginning of the week. In the first period, questions posed were centered on the lecture note to exemplify and to analyze it. In the second period, questions posed were centered on video clips or episode dialogs to allow pre-service teachers to focus on teaching processes, with the aim of improving students’ learning. For example, for the first chapter called “Introduction to Course and Effective Teacher”, the opened threads were: (1) some terms used in the course, (2) effective teacher, (3) teacher’s level of subject matter, and (4) psychological characteristics. Examples of questions posed for the second thread “effective teacher” in the first and second period of the week were: “Is this statement true: “Successful students will be effective teachers?” and “Is the teacher in the video clips an effective one?” respectively. In the last period, they prepared the summary of the forum discussions in the light of the lecture notes and discussions.

In WBIIG, lecture note related to each week, in contrast to WBDI, was not provided to the pre-service teachers at the beginning of the week. In other words, the content in WBDI was clearly on front of the pre-service teachers. They elaborated and discussed the content, and attempted to analyze examples. Questions in WBDI require restating the information in the lecture note whereas questions in WBII require sharing of experience and ideas. Instructor in WBDIG did not need to use encouraging prompts to provide collaboration between pre-service teachers on the content and pre-service teachers did not need help to acquire the necessary knowledge. On the other hand, pre-service teachers in WBIIG discussed guiding questions and they continuously needed help and collaboration to construct knowledge. WBII questions posed in the first period provided opportunities for learners to collaboratively
construct knowledge based on discussions and reflection, and to articulate and revise their thinking in order to ensure the accuracy of knowledge construction. The questions posed in the second period were centered on video clips or episode dialogs to allow pre-service teachers to focus on teaching processes, with the aim of improving students’ learning. For instance, for the first chapter called “Introduction to Course and Effective Teacher”, the same four threads were opened as in the WBDIG. Examples of questions posed for the second thread “effective teacher” in the first and second period were: “Recall both effective and ineffective teachers you may have had! To what extent did they seem to differ with respect to the knowledge of how to teach the subject, or knowledge of how people learn?” and “What would you do if you were in place of the teacher in the video clips?” Why or why not? In the last period, they constructed the lecture notes based on the forum discussions and research done through the periods on the Internet. Lastly, lecture notes prepared beforehand by the researchers were also opened at the end of the week.

Figure 1. WBDI Course Flow for Each Week
In both groups, each pre-service teacher was required to participate at least three times to the discussions in total for the first two periods in each week. Each message of the pre-service teachers in the forum discussions was assessed by using the coding technique developed by McKinnon (2000). After coding the messages in the forum discussions, the instructor used a grading rubric developed by the researchers to score the pre-service teachers’ messages under these codes. The discussion scores of the pre-service teachers related to each week were announced on Sunday at that week. These scores constituted 25% of the course grade.

**Power Analysis**

An essential and primary decision in the power analysis is the determination of the population effect size before the study. Cohen and Cohen (1983) offered the following values; small, ES = .20; medium, ES = .50; and large ES = .80 [ES = (Mean of the WBII group – Mean of the WBDI group / standard deviation of WBDI group)]. As many studies
proposed, constructivist approach is appropriate to the nature of web-based learning environments (Jonassen, 2000; Keagan, 1997; Willis, 1995). Therefore, effect size was set to high ($f^2 = 0.33$ for variance and 0.8 for mean difference) at the beginning of the study. During the analyses, the probability of rejecting true null hypothesis (probability of making type-1 error) was set to .05 as a priori to our hypothesis testing, as it is the mostly used value in educational studies. This study was conducted with 63 pre-service teachers considering nine variables. Then the power for that sample size and large effect size was calculated for nine variables. Power of this study was calculated as .85. Therefore, the probability of failing to reject the false null hypothesis (probability of making Type 2-error) was found as .15.

**Results**

Only one of the 63 pre-service teachers did not return PREGMQ. Missing PREGMQ data constituted 1.5 % of the whole data. Since this missing data constituted a range less than 5 % of the whole data, it was directly replaced with the series mean of the entire subjects as suggested by Cohen and Cohen (1983).

**Determination of Covariates**

Six covariates (age, program, gender, GPA, PRECFW and PREGMQ) were predetermined as potential confounding factors prior to conducting the MANCOVA. In order to limit the variables in the covariate set to a few reliably measured and highly correlated ones with the dependent variables (Cohen & Cohen, 1983), these potential covariates were correlated with the dependent variables (see Table 1). Gender, GPA, PRECFW and PREGMQ had significant correlations with at least one dependent variable. Therefore, they remained in the covariate set for the inferential statistics and the rest were discarded.

**Homogeneity of Regression**

To check the homogeneity of regression assumption, Multivariate Regression and Correlation (MRC) was conducted using enters method for each variable to test the significance of $R^2$ change for four interaction terms produced by multiplying the group membership with the covariates, separately. The contribution of interactions is not significant for the ME, FE, WBAS, and POSTGMQ [$F(6, 49) = 1.768, p=.125$, $F(6, 49) = .186, p=.980$, $F(5, 51) = .673, p=.672$, and $F(6, 49) = 1.074, p=.391$, respectively]. These results indicate that there are no significant interactions between covariates and the group membership; therefore the interaction set was excluded from further inferential statistical analyses.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>ME</th>
<th>FE</th>
<th>WBAS</th>
<th>POSTGMQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-.154</td>
<td>-.143</td>
<td>-.157</td>
<td>-.021</td>
</tr>
<tr>
<td>Program</td>
<td>.059</td>
<td>.156</td>
<td>-.051</td>
<td>-.180</td>
</tr>
<tr>
<td>Gender</td>
<td>-.274*</td>
<td>-.305*</td>
<td>.014</td>
<td>.088</td>
</tr>
<tr>
<td>GPA</td>
<td>.450*</td>
<td>.391*</td>
<td>.117</td>
<td>.299*</td>
</tr>
<tr>
<td>PRECFW</td>
<td>-.026</td>
<td>.086</td>
<td>-.292*</td>
<td>-.219</td>
</tr>
<tr>
<td>PREGMQ</td>
<td>.377*</td>
<td>.282*</td>
<td>-.264*</td>
<td>.702*</td>
</tr>
</tbody>
</table>

* Correlation is significant at the .05 level (2-tailed)

**The MANCOVA Model**

A MANCOVA model conducted revealed that there is a significant main effect for instructional methods (treatment), Wilk’s $\lambda = .509$; $F(6, 50) = 8.043$; $p < .05$ (See Table 2). Subsequent to the significant omnibus MANCOVA presented above, protected univariate F tests and step-down analysis were conducted to investigate the unique importance of each dependent variable in the model.
Protected Univariate F Tests: A univariate analysis of covariance (ANCOVA) conducted as follow-up tests of the MANCOVA indicated a statistically significant mean difference between groups for the FE in the favor of WBIIG \(F(1, 55) = 31.725, p < .05\) and for the WBAS in favor of the WBDIG \(F(1, 55) = 8.412, p < .05\). On the other hand, there is no statistically significant difference for the ME and the POSTGMQ \(F(1, 55) = .080; p > .05, F(1, 55) = .011; p > .05\), respectively. This means that pre-service teachers taught by the WBII got higher scores on the final exam than the pre-service teachers instructed by the WBDI. Moreover, the pre-service teachers in the WBDIG developed a more positive attitude towards web-based course than the pre-service teachers in the WBIIG. The descriptive statistics for each dependent variable are reported in Table 3.

Table 2. Multivariate Tests Results for the MANCOVA

<table>
<thead>
<tr>
<th>Effect</th>
<th>Wilks’ Lambda</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.880</td>
<td>1.138</td>
<td>6</td>
<td>50</td>
<td>.354</td>
<td>.120</td>
<td>.406</td>
</tr>
<tr>
<td>AGE</td>
<td>.883</td>
<td>1.105</td>
<td>6</td>
<td>50</td>
<td>.400</td>
<td>.092</td>
<td>.394</td>
</tr>
<tr>
<td>GENDER</td>
<td>.840</td>
<td>1.588</td>
<td>6</td>
<td>50</td>
<td>.400</td>
<td>.160</td>
<td>.556</td>
</tr>
<tr>
<td>GPA</td>
<td>.778</td>
<td>2.379</td>
<td>6</td>
<td>50</td>
<td>.400</td>
<td>.222</td>
<td>.763</td>
</tr>
<tr>
<td>PRECFW</td>
<td>.928</td>
<td>.644</td>
<td>6</td>
<td>50</td>
<td>.400</td>
<td>.072</td>
<td>.233</td>
</tr>
<tr>
<td>PREGMQ</td>
<td>.569</td>
<td>6.320</td>
<td>6</td>
<td>50</td>
<td>.400</td>
<td>.431</td>
<td>.997</td>
</tr>
<tr>
<td>TREATMENT</td>
<td>.509</td>
<td>8.043</td>
<td>6</td>
<td>50</td>
<td>.400</td>
<td>.491</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 3. Descriptive Statistics of the Dependent Variables

<table>
<thead>
<tr>
<th></th>
<th>WBIIG</th>
<th>WBDIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>60.16</td>
<td>59.15</td>
</tr>
<tr>
<td>FE</td>
<td>67.48</td>
<td>54.85</td>
</tr>
<tr>
<td>WBAS</td>
<td>111.72</td>
<td>129.03</td>
</tr>
<tr>
<td>POSTGMQ</td>
<td>101.03</td>
<td>101.71</td>
</tr>
</tbody>
</table>

Step-down Analyses: As there was a significant correlation between the FE and the WBAS \(r = .01, p < .05\), there was a need to proceed a step-down analysis for the FE as the dependent variable of highest priority, and the WBAS as the additional covariate. The results showed that the method of instruction had significant effect on the FE \(F(1.54) = 35.925, p = .000\). This implies that after accounting its effect on WBAS, the effect of instructional method on pre-service teachers’ final exam is still significant.

Treatment Verification

The instructor followed all of the forum activities, emails, and other web tools and took notes based on the OC criteria about the flow of the lesson. Through the treatment period, he completed OC for six weeks (six chapters) in both groups by using the notes taken during the implementation. The instructor-completed observation was compared by those completed by the outside observers (one experienced teacher for each group) and the interrater reliabilities for the WBII and WBDI were .92 and .88, respectively.

To compare the two instructional methods, the six observations of the instructor for the two courses were used. For this purpose, t-test was conducted. The results revealed that there was significant difference between two instructions on each criterion of the OC except three criteria. These three criteria are related to online settings, which are valid for any proper web-based learning environment. So, it means that the two courses were administered significantly different from each other on the base of the instructional methods. According to these observation scores for each group and the implication from the t-test, it could be assumed that teaching methods course in both groups was conducted according to the treatment requirements.

Conclusion and Discussion

The main purpose of this study was to investigate the effects of the asynchronous web-based “General Teaching Methods” courses conducted based on direct or indirect instructional methods on pre-service teachers’ achievement, metacognition, and attitudes towards web-based course. The following conclusions are offered: (a) WBI was more
effective on the pre-service teachers’ achievement than WBDI did, (b) WBDI was more effective on pre-service teachers’ attitude towards WBC than WBII did, and (c) there was no significant difference between the WBIIG and WBDIG on metacognition.

The observed value of effect size was calculated for each dependent variable where significant difference was seen. The calculated eta squared was .512 for the total model, .383 for the FE and .142 for the WBAS. Olson and Wisher (2002) reported the results of the 15 studies between the years 1996 and 2002 about the effects of the web-based instruction on the students’ outcomes, such as learning, performance, and satisfaction. In their analysis, effect sizes ranged from -.16 to .40. The grand mean for all 15-effect sizes was .24. So, this study supported previous researchers’ findings. The effect sizes for the total model and the FE are large, whereas the effect size for the WBAS is medium. In other words, results for the FE and for the total model were of practical significance for similar populations.

The significant group differences on the FE support the hypothesis that WBII promotes the achievement. This result also replicate the achievement benefits in mathematics (Sorg, 2000), and educational courses (Goggin, Finkenberg, & Morrow, 1997; Johnson, 2001; Matuga, 2001; Pevato, 2003; Tsai, 2004) associated with WBII that has been observed in the undergraduate and graduate. Insignificant difference between groups on the ME is consistent with the findings of the previous studies (Chellman & Duchastel, 2000; Cyrs & Conway, 1997), which reported that the online learning environment has a novice effect on the students if they take the web-based course for the first time. Due to novice effect, the effect of the treatment remains in secondary level at the beginning of the class. As a matter of fact, novice effect was, especially, seen in the discussion forum. Participation in forum discussions and cohesiveness between threads for both groups improved with time. They investigated the links in the course site more, searched Internet more and sent rich and deep messages containing critical and creative thinking, and intensive interactions. Mean of the discussion scores increased from 5.2 to 7.8 in the WBIIG and from 5.6 to 6.4 in the WBDIG through the course. Similarly, mean of the interaction scores for the WBDIG were increased from 3.2 to 6.8 and that for the WBIIG were increased from 3.5 to 7.1 through the course. This might culminate in the high achievement in the final exam.

Another explanation for the question why WBIIG compared to WBDIG produced a higher performance on the FE might be that WBII required an active process within which students built on previous knowledge and developed personal connections to conceptual material. Indirect instruction paradigm reflects a position that knowledge is not independent of the learner but is internally constructed by the learner as a way of making meaning of experiences. As a result of that the facilitative questions in the WBII focused on the students’ experiences, readings about the concepts and exploring ill-structured cases (videos or episodes) rather than on the definitions and concepts in the lecture notes as in the WBDI. Questions on experiences and ill-structured cases were the heart of the forum discussions in WBIIG to facilitate discovering the concepts of the topic and pre-service teachers presented their ideas by focusing on them. WBII is a developmental process beginning from duality, moving to an understanding of multiple views, and finally acknowledging the context wherein the solution is given (Jonassen, Davidson, Collins, Campbell, & Haag, 1995). Therefore, although the same learning materials were used in both groups, the content of the questions as well as flow of instructions stimulated pre-service teachers differently.

The study found that there is a significant difference between WBIIG and WBDIG in the attitude towards web-based course. Consistent with previous studies (Manuel, 2001; Matuga, 2001), reaching the content easily and obtaining knowledge by spending less effort can explain the difference in attitude towards web-based course in favor of WBDI. The relatively low attitude towards web-based course of WBIIG might be caused by the need of more effort to acquire the content of the course than of the WBDI and it culminated in the need of more time for the familiarity. Familiarity with WBC, in general, was highly correlated with initial attitude towards WBC (Richardson & Price, 2003). As pre-service teachers had no previous experience about the web-based learning environment, the high amount of time-on-task about the challenging subjects caused frustrations. It is similar to the previous researchers’ reports (Chellman & Duchastel, 2000; Cyrs & Conway, 1997). This finding was also supported by the WBIIG’s messages in the informal forum discussion of the course, such as “I studied this course as three times as any other course” or “If I study for any one of my course as much as for this course, I can certainly take a credit AA.” But, this high effort to acquire the content, in turn, produced high performance in the FE for WBIIG. By providing basic facts and principles without spoiling the flow of discussions in a timely manner in WBII, instructor or moderator might decrease cognitive loads on students. Depending on the students’ attitude, at the beginning lessons, WBDI can be conducted. Later, WBII can be conducted.
Another important finding of this study showed that there was no significant difference between the two groups with respect to general metacognition. It might be caused by the short period of the teaching/learning process or by the pre-service teachers’ high metacognition. Many studies on metacognition reported that they need longitudinal research to observe the change and development in the metacognitive strategies depending on training on them (Schraw & Moshman, 1995; Schraw, 2000). Consequently, activities or tasks might be embedded into online projects, readings, and threads in the online discussions in order to train students in terms of metacognitive abilities. The metacognition of the pre-service teachers in both groups were moderately high or high (only three for WBDIG and two for WBIIG were low). Therefore, future research is needed to be conducted in a more representative sample, which includes students in every metacognition levels. Additionally, research should also attempt to determine differences in the achievement and attitudes of the learners with high and low levels of metacognitive abilities.

There were several limitations to the findings of this study that merit attention. First, although the sample size provided sufficient power for the desired statistical analyses, it was smaller than ideal. Future studies could address this issue through larger and more representative samples and ideally using true experimental design. Although the experimental and control group was randomly assigned, the participants in both groups were administratively defined (already existed group).

An additional limitation comes from the self-report nature of the GMQ and the WBAS. Self-report measure of metacognition does not guarantee that the participants actually had and used the metacognitive abilities. Self-report measure of attitudes does not guarantee that the participants felt this way all through the course. Behavioral measures or think-aloud protocols during the treatment would have enabled more certainty as to whether participants were actually using metacognitive abilities and how they felt during the course of the experiment.

Another limitation of this study concerns the dealing of pre-service teacher belong to two different programs. This created difficulty focusing on deeply in mathematics and science. Additional research is needed to fill the gaps in our understanding of the pre-service teachers’ achievement and attitudes during the Asynchronous Web-Based Mathematics or Science Method Course. In this respect, scores obtained from online discussion forum should also be used to determine course achievement besides paper-based exams. It might positively affect both achievement and attitude towards web-based course.

The implication of this study for pre-service teachers involves the importance of indirect instruction on academic performance but not much on attitude.

References


Appendix

Example Questions from the Midterm Exam

1. “Who can tell me the cell division?” is a leading question.
   True                                           False

2. Focusing the subject on one idea at a time and presenting it so that learners can master a point before the teacher introduces the next point is the key of the ________________ method.

3. Directions: Match the description in column A with the correct method in column B. Write the matching letter in the blanks provided in column A. The items in column B may be used once, more than once, or not at all. Consider the primary characteristics of the method!

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teaching how to deal with new situation</td>
<td>A. Questioning</td>
</tr>
<tr>
<td>2. Helping students to evaluate their own progress</td>
<td>B. Problem Solving</td>
</tr>
<tr>
<td>3. Providing students a structured presentation.</td>
<td>C. Discovery</td>
</tr>
<tr>
<td>4. Helping students to organize and build.</td>
<td>D. Lecturing</td>
</tr>
<tr>
<td>5. Helping students to monitor and regulate their action.</td>
<td>E. Expository</td>
</tr>
<tr>
<td>6. Helping students to recognize their environment</td>
<td></td>
</tr>
</tbody>
</table>

4. Which one the following is NOT the primary limitations of the lecturing (a type of direct instruction)?
   A. Audience is passive   B. Learning is difficult to gauge
   C. Communication in one way   D. Time used for lecturing

5. First “Express 3/8 as a decimal” then draw the distinctions in the solution related to the different types of mathematical competence (Fact, skill, concept, strategy) possessed by an individual. (8 points)

Example Questions from the Final Exam

1. Project-based learning is an instructional method that aims at student engagement in real world tasks to enhance learning.
   True                                         False

2. The question “How do single-celled animals reproduce themselves and divide up to create similar animal life that looks like themselves?” is an ineffective question because it is ________________-question. It should be stated as: ________________

3. Directions: In the following exercise, column A contains description of students’ activities in team-oriented cooperative learning and column B names of three types of the team-oriented cooperative learning. Match each activity in column A with its corresponding type of team-oriented cooperative learning. Write the matching letter in the blanks provided in column A. Team-oriented cooperative learning types may be used one, or more than once.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Students learn the content through the teacher presentation.</td>
<td>A. Students Teams - Achievement Division</td>
</tr>
<tr>
<td>2. Students read the unique section of the text at the beginning of the class.</td>
<td>B. Jigsaw</td>
</tr>
<tr>
<td>3. Students monitor teammate’s level of understanding.</td>
<td>C. Team-Assisted Individualization</td>
</tr>
<tr>
<td>4. Students study through the assigned unit given by the student’s monitor.</td>
<td></td>
</tr>
<tr>
<td>5. Students take worksheet and work on it with their team.</td>
<td></td>
</tr>
<tr>
<td>6. Expert groups help the students by giving information about the content.</td>
<td></td>
</tr>
</tbody>
</table>

4. Which one of the following is NOT TRUE for the drama as a method?
   A) Experiential approach is used for learning
   B) Product is the major goal of the drama.
C) A variety of communication experiences are offered.
D) Teacher becomes a learner among the learners

5. Read the case below! Assume that the learning task is the organelles of an animal cell and the subject is explained by the drama. Each of the organelle is matched with a student.
Student N- Nucleus
Student M- Mitochondria
Student G- Golgi apparatus

The students stand in the center of the class. Firstly, students introduce themselves. For instance; Student R says:” My name is ribosome. I’m the smallest cell and I’m responsible for the protein synthesis in the cell.” After students introduce themselves, they act the activities of the cell while it digests a harmful bacterium. The students show the relations between the organelles and display the functions.
Analyze the concepts below by using the drama example above.

Dramatic enactment:
Dramatic moments (At least two)
Communication forms (at least two)
Representation forms (at least two)
Building a Model Explaining the Social Nature of Online Learning

I-Chun Tsai, Bosung Kim, Pei-Ju Liu, Sean P. Goggins, Christiana Kumalasari and James M. Laffey

School of Information Science and Learning Technologies (SISLT), University of Missouri-Columbia, USA // itch9@mizzou.edu // bkq22@mizzou.edu // pl60t@mizzou.edu // outdoors@acm.org // ckelf7@mizzou.edu // laffeyj@missouri.edu

ABSTRACT

Based on a framework emphasizing the social nature of learning, this research examines a model of how social constructs affect satisfaction within online learning using path analysis for students in higher education. The social constructs evaluated in this study include sense of community (SOC), social ability (SA), perceived ease of use (PEU) and usefulness (PU) of social awareness tools, and self-reported participation (SRP) in online courses. Students’ social ability is predicted by perceived ease of use of social awareness tools and self-reported participation. Satisfaction with online learning is explained by sense of community, social ability, and self-reported participation. Sense of community is shown to be a mediator of the relationships between social ability and satisfaction (SA→S) and between perceived usefulness and satisfaction (PU→S). Additionally, students’ social ability, perception of usefulness of social awareness tools, and self-reported participation serve as predictors for students’ sense of community. The findings from the path analysis show the value of simultaneously examining the relationships among the multiple social constructs, rather than simply examining relationships of how one construct may predict another as has been shown in previous studies.

Keywords

Sense of community, Social ability, Technology Acceptance, Participation, Online learning satisfaction, Path analysis

Introduction

Online learning has become a pervasive part of higher education. Online enrollment increased from 1.98 million in 2003 to 2.35 million in 2004, and approximately 74 percent of public institutions of higher education identified online education as a critical long-term strategy in 2005 (Allen & Seaman, 2005). Many positive reports of online learning success show its impact and potential, such as relative equivalence in test-result outcomes with face-to-face courses (Talent-Runnels, et al., 2006), broad implementation and rapid growth across higher education, and provision of access for many students who would otherwise have to forego higher education. Although students appreciate the flexibility and convenience offered by online learning environments, online students do experience a sense of isolation (Abrahamson, 1998; Bessar & Donahue, 1996; Rahm & Reed, 1998), and express being more satisfied with face-to-face courses (Allen, Bourhis, Burrell, & Marbry, 2002; Simonsen, 1997; Klesius, Homan, & Thompson, 1997). In her study, Carr (2000) found higher dropout rates for distance education (10-20%) over traditional programs. Reasons given for the high dropout percentage of distance learners include limited support and service of distance education, dissatisfaction with teaching methods, unfamiliarity with the technology used, and student feelings of isolation. Hara and Kling (2000) also found that online students were frustrated by the communication and technical difficulties that impeded interaction. Arbaugh (2000) argues that the lack of social interaction was a factor that depressed student satisfaction in online learning. This dissatisfaction with online learning can be seen in high rates of attrition for online students (Chyung, 2001). Ashar and Skenes (1993) found that while adults were attracted to a higher education business program because of strong learning needs; those needs were not strong enough to retain them. However, retention was positively impacted by establishing a social environment within the program. Students need to feel involved and develop relationships with other students in an online course (Rovai, 2002a). Following from an appreciation of the social nature of learning, learning and cognitive development are recognized as substantially constituted through social participation and interaction (Vygotsky, 1978; Wenger, 1996). In order to understand how students’ online learning satisfaction is affected by their social participation and interaction with others, the present study seeks to build a model of the elements that contribute to the social nature of online experience and influence satisfaction in online learning environments. We begin by framing our research with key social constructs for understanding the social nature of online learning. Next, we describe the use of path analysis, to examine the relationships among these constructs. Our results, present a model for explaining how these social constructs influence satisfaction in online courses.
Theoretical Perspectives

Advances in our understanding of how we learn show that context play a significant role in determining what we learn and how we will be able to use what we learn. Wenger (1998) describes social participation as a process for learning and knowing which is relevant “not just to local events of engagement in certain activities with certain people but to a more encompassing process of being active participants in the practices of social communities and constructing identities in relation to these communities” (p.4). In a community of practice, people learn by socially negotiating meaning of the world with what they see, who they know, and what they do. Through negotiation about practice people expand what they know and are able to do, as well as learn from others’ actions and feedback. Their growth of knowledge is represented not only in individual change but also in the shared values, relationships, networks, and knowledge produced when interacting with others. To be more specific, individual identity interacts with cognition and social relationships in a community of practice. Community practices influence individual identity, and individual identity and cognition contribute to the negotiation of new views of the community.

The social nature of educational practice influences students’ motivation to learn, ways of participating, negotiation of meanings for new knowledge and skills, and how new learning shapes self-identity and community membership. Online learning environments are social contexts, just as are face to face environments. However, these social activities and contacts are mediated by technology tools within a specific networked environment. Researchers studying online learning environments explore the social nature of learning in a variety of ways, such as students’ use of communication tools (Hara & Kling, 2000), students’ feeling of isolation (Abrahamson, 1998), students’ perception of a sense of presence (Picciano, 2002), and the relationship between students’ feeling a sense of belonging and their amount of social interaction (Rovai, 2002c). Little is known, however, about how usage of communication tools, sense of presence, and interaction are interconnected in the way they influence isolation, satisfaction, or other student outcomes. New knowledge is needed to understand how students experience the social aspects of their online courses, and how the tools and methods of online learning can foster sociality and social interaction. To explicate the social nature of online learning and how it influences students’ learning satisfaction, we examine four constructs of online social life: sense of community, social ability, technology acceptance, and participation.

Sense of community

McMillan and Chavis (1986) define sense of community as “a feeling that members have of belonging, a feeling that members matter to one another and to the group, and shared faith that members’ needs will be met through their commitment to be together” (p.9). Lave and Wenger (1991) argue “activities, tasks, functions, and understandings do not exist in isolation; they are part of broader systems of relations in which they have meaning. These systems of relations arise out of and are reproduced and developed within social communities, which are in part systems of relations among persons” (p. 53). Wenger implies the interrelationship between sense of community and social interaction when he describes the integration of community membership with personal identity. Researchers report a variety of benefits derived from a strong sense of community. Rovai (2001) found that the sense of classroom community was moderately and positively related to the number of messages posted to the discussion boards. In turn, the comments posted by learners provide evidence that more interaction and participation promoted sense of community in class. This increased sense of classroom community was attributed to how interaction and involvement lessened psychological distance of online students in the course activities. Rovai’s (2001) findings are aligned with previous research that found students with stronger sense of community tend to have a greater flow of information exchange with other members, have higher availability of support, and have a higher commitment and satisfaction to group work and collaboration (Wellman, 1999; Dede, 1996; Bruffee, 1993) as well as a greater sense of well being (Rovai, 2002a). Specifically, Tinto (1993) and Scott (2004) showed that students’ satisfaction with online courses increases when they feel involved and have strong relationships with members in the learning community. Sense of community has also been positively associated with students’ learning achievement, although these results were not from learning in online community (Bryk & Driscoll, 1988).

Sense of community is recognized as an important factor for fostering interactivity or interaction among participants in online learning environments; however, building and sustaining a sense of community in a virtual classroom is challenging. According to Rovai (2002a), the key for building and sustaining a sense of community is facilitating interaction among students. He observed students reported better sense of community in class when they felt higher
levels of interaction. Then, what can the online learning community do to foster interactivity or interaction among participants in online learning environments so that students can build a positive sense of community? Two types of interactions: task-driven interactions and socio-emotional interactions have been identified from previous study. Rovai (2001) argues that task-driven interactions may not be enough to foster a positive sense of community among participants and a supportive environment for socio-emotional interactions is critical for promoting a sense of community. Rovai (2002) found that helping students be aware of other members’ activities is one way to support students’ sense of community.

Social ability

Social ability represents how able members are in using the resources of the social context to achieve important goals (Laffey, Lin, & Lin, 2006). Social ability in online learning environments is determined by the fit among people, tools, and activities. For example an instructor may be able to work with a project group in a face-to-face classroom, but if the task is mediated by a chat tool which is unfamiliar or awkward for the instructor, the group may not advance as effectively as it could in a face-to-face context due to the poor fit between the instructor and the chat tool. Through several exploratory studies with online students, five components of social ability were explicatied: written communication skills, peer social presence, instructor social presence, comfort with sharing personal information, and social navigation (Laffey, Lin, & Lin, 2006; Yang, Tsai, Cho, Kim, & Laffey, 2006; Lin, et al., 2006). In other words, students who have higher social ability in online courses tend to be more aware of and act upon peer and instructor’s actions, to share personal information, and to communicate effectively with others in the written format.

These capabilities have been shown to impact students’ online learning processes and outcomes. In particular, written communication skills matter! Writing skills are a necessity in online environments, since online synchronous and asynchronous communication channels remain primarily text-based, especially in the most common online course tools like discussion board, email and chat. The ability to clearly communicate through writing has been recognized as a critical requirement and predictor of student success in online courses (Golladay, Prybutok, & Huff, 2000; Mandernach, Donnelli, & Dailey-Hebert, 2006). Students who have poor writing skills may experience difficulties expressing opinions and building relationship in online courses. In his observation of 20 graduate students, Rovai (2001) found that students’ communication styles influence building a sense of community. Students who wrote a message using a supportive, helpful, and connected voice were more likely to feel a stronger sense of community than those with an impersonal, assertive, and independent voice. Since there are limited verbal exchanges in online courses, student learning and interaction heavily rely on written products of communication. More specifically, clearly communicating questions, sharing opinions, expressing concerns, and providing support in written forms are abilities that influence the social nature of online learning (Mandernach, et al, 2006).

Additionally, Garrison and Anderson (2003) argued that the formation of community requires a sense of social presence among members in the online environment. Social presence is defined as “an attribute of computer-mediated activity, derived from media studies about how effectively media (TV, etc.) convey the sense that mediated participants were really present (Short, Williams, Christie, 1976)” (Laffey et al., 2006, p.166). Reidl (2001) also stated that social presence was the degree of salience of another person in a social interaction, while Wise, Chang, Duffy and del Valle (2004) characterized social presence as the feeling that others are participating with you. Yang et al. (2006) found that students perceived peer and the instructor presence differently. Further, social ability also represents the capacity to connect and keep members engaged in interaction. Sharing personal information has been shown to foster the formation of sense of community by facilitating deeper and more meaningful interpersonal relationships in online communities (Swinth, Farham, & Davis, 2002). Moreover, the notion of social navigation defined as “a construct representing being aware of what others are doing as a primary guide for one’s own action” (Laffey et al., 2006, p.166) means using information about the action of others to guide action and decision making. Thus, the capability of students to observe the “footprints” of others before making decisions has the potential to impact the efficiency and effectiveness of online learning and contribute to a greater sense of community.

Along with sense of community, previous research suggests that social ability is a key determinant of how students experience online learning. Lin et al. (2006) showed that social ability was a significant predictor for students’ online learning satisfaction and was a critical construct for explaining students’ changes from a peripheral into a central role in a community. Also, students’ participation and interaction is somewhat determined by how well the tools scaffold their social needs and achievement of learning goals. Social ability as an indicator of the fit among person, task and tool is highly affected by the affordances provided by technology in a context.
Technology acceptance

Previous studies recognized that students’ familiarity with the technology use and feeling of how they are supported by the communication and interaction tools influenced their satisfaction with online learning (Carr, 2000; Hara & Kling, 2000). Also, Gallini and Helman (1995) found that online learners need to be effective in communicating information particularly information related to others’ actions in order to learn successfully. Since technology tools play a deterministic role in how well students can interact with others their acceptance by students influences satisfaction with online interaction and learning. The Technology Acceptance Model (TAM) has been one of the most influential models in predicting and explaining usage and acceptance of a new technology. According to Davis (1989), perceived usefulness and perceived ease of use are the two major factors affecting users’ acceptance behaviors. Davis (1989) defines perceived usefulness as “the degree to which an individual believes that using a particular system would enhance his or her job performance” (p.320) and perceived ease of use as “the degree to which an individual believes that using a particular system would be free from physical and mental effort” (p.320). Davis argues that users accept and use a technology primarily because of the functions it performs for them, and secondarily because of how easily they can get the technology to perform the tasks.

In online environments, the use of technology tools is necessary for social interaction and subsequent learning. One of the key challenges of using technology in social interaction is awareness of who is available, what has happened, and what is going on. Without special design provisions using course management system can leave the student feeling socially blind and isolated. Dourish and Bly (1992) examined how synchronous and asynchronous computer mediated communication (CMC) tools deliver and support social awareness. They found successful delivery of social awareness via the CMC tools resulted in a sense of community through maintaining working relationships and informal communication. Also, Schwier (2002) indicated that accessibility and use of technology tools for online courses is critical to facilitating and building virtual learning community because technology tools provide a gathering space and mediated communication channel for students. Adapting TAM to examine students’ online learning behavior, Lin (2005) found that students’ intention to use technologies affected their usage behavior in online learning environments.

Participation

Wenger (1998) defines participation as “the social experience of living in the world in terms of membership in social communities and active involvement in social enterprises” (p.55). Wenger also states that “Knowledge, belonging, and doing are not separable: What we know, who we are and what we do seamlessly come together in one experience of participation” (Wenger, 1996, p.22). The learning process can be described as how students move from peripheral participants toward becoming core participants in a learning community. This transformation in the way they participate is also represented in how they perceive themselves in the learning community, and how they are perceived by others (Wenger, 1998). Thus, participation, learning and membership in a community are intertwined, and members’ identities change through their participation. Since learning is constituted through social activities, sense of community, which describes students’ sense of belonging and a social bond with others, has potential to critically and positively influence learner participation and interaction (Rovai, 2003; Carroll, 2001; Putnam, 2000). Reciprocally, students’ active participation is often seen as the critical element in the formation of a learning community and building a sense of community in online classes (Wang, Sierra, & Folger, 2003).

Researchers have identified several methods to encourage student participation and interaction in online learning, such as instructional design for collaborative course activities (e.g. group work, debate, web polls, etc.), and the use of different types of synchronous and asynchronous communication tools (e.g. chat, discussion board, blog, etc) (Collins & Zane, 1996). Further, a qualitative study by Falvo and Solloway (2004) found that the online learning format, technology used for supporting learning, instructional design, and various social activities and relationships contribute to a sense of community recognized to be influential for student participation. They showed that explicit design of the learning environment to encourage students to identify with the course as a learning community both fostered discussion about what it means to be an online community, and an actual sense of community. Participants reported satisfaction with the format of online learning because of the advantages of learning anytime and anywhere and felt interested in using different types of technology for interaction among students and the instructor. Active participation through usage of the technology tools seems to be crucial for enabling learners to accomplish learning.
tasks. How well learners appropriate the tools to socially interact with others or participate in activities and how they feel about themselves and others provide a basis for inviting and reinforcing participation.

Research Questions

In summary, previous literature about online learning has shown that failure to achieve a sense of community and feelings of isolation negatively affect acceptance of and satisfaction with online learning (Vonderwell, 2003; Woods, 2002). The current study seeks to test and build a model of social factors that are hypothesized to influence sense of community and further to see if this model also explains satisfaction with online learning. Path analysis is used to determine the extent to which key constructs, including self-reported participation, social ability, perceived ease of use and usefulness of social awareness tools, and sense of community explain satisfaction with online learning. The research questions of this study are:

(1) To what extent does sense of community influence students’ overall learning satisfaction in online learning?

(2) How well does the final path model for social ability, self-reported participation, and technology acceptance explain sense of community?

Research Method

Participants

Near the end of Spring semester 2006, recruiting emails with links to an electronic consent form and the survey instruments were sent to 82 students who were enrolled in five online courses. The online courses mainly targeted learning to use technologies or learning to design learning systems. On average, each course has about 16 to 17 students. There was a 63.4% response rate, resulting in 52 participants who filled out the survey over the Internet. After initial data screening, one case was eliminated as a univariate outlier. Table 1 presents the demographic information for the 51 subjects.

<table>
<thead>
<tr>
<th>Demographic Information</th>
<th>Number of Participants</th>
<th>Percentage (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23</td>
<td>45.1</td>
<td>51</td>
</tr>
<tr>
<td>Female</td>
<td>28</td>
<td>54.9</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native Speaker</td>
<td>40</td>
<td>78.4</td>
<td>51</td>
</tr>
<tr>
<td>Non-native Speaker</td>
<td>11</td>
<td>21.6</td>
<td></td>
</tr>
<tr>
<td>Academic Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>4</td>
<td>7.8</td>
<td>51</td>
</tr>
<tr>
<td>Graduate</td>
<td>47</td>
<td>92.2</td>
<td></td>
</tr>
<tr>
<td>Previous Online Courses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1 courses</td>
<td>9</td>
<td>17.6</td>
<td>51</td>
</tr>
<tr>
<td>2-5 courses</td>
<td>17</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>&gt; 6 courses</td>
<td>25</td>
<td>49.0</td>
<td></td>
</tr>
<tr>
<td>Hours Login(weekly)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5 hr.</td>
<td>20</td>
<td>39.3</td>
<td>51</td>
</tr>
<tr>
<td>6-10 hr.</td>
<td>17</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>&gt; 10 hr.</td>
<td>14</td>
<td>27.5</td>
<td></td>
</tr>
</tbody>
</table>

Context

The data were collected from five online courses offered in a college of education in a mid-west university during Spring semester 2006. All five courses were delivered fully online through the Sakai 2.0 course management system. These five courses had similar course structures with a typical unit comprising a set of learning tasks that directed students to work individually or interactively with peers to accomplish assignments. Besides the course management system, a daily email digest (Appendix A) of course activity was implemented using CANS (Amelung, 2005). This email digest was a daily summary with detailed information about all the previous day’s activities within the course site, and it was automatically sent out to all students at mid-night so that students received it each morning.
actual content of an email digest is a list of events for each of the main tools, such as discussion board, chat room, resources, announcement, and assignment tools. The purpose of providing the email digest for students is to support their social awareness of what others were doing in the course. Instead of needing to enter the course management system and browse through all of the site tools, reading the digest could inform the student that many students posted messages in a thread or uploaded files the day before. This sort of information could remind students about work to be completed or make them curious about what others were saying or doing.

**Instruments**

The instrument applied in this study is adapted from several instruments measuring the social constructs addressed in this study. Below is the explanation of how the items for measuring social constructs are adapted. The survey items of these social constructs applied in this study are presented in Appendix B.

**Sense of Community.** Sense of community was measured with 20 items using a 7-point Likert scale from the Classroom Community Scale (Rovai, 2002b). The Cronbach $\alpha$ reliability estimates from our data were .93 for the sense of community.

**Social ability.** An Online Learning Experience Study Questionnaire (OLESQ; Yang et al., 2006) consisting of 30 items that measure students’ perceived social ability in online learning environments was used in this study. Results from the OLESQ show that five factors accounted for 61.86% of the variance in the measure: perceived peer social presence (10 items, $\alpha = .93$), perceived written communication skills (3 items, $\alpha = .90$), perceived instructor social presence (8 items, $\alpha = .910$), comfort with sharing personal information (3 items, $\alpha = .83$) and social navigation (6 items, $\alpha = .88$). The Cronbach $\alpha$ reliability estimates from our data were .92 (30 items) for social ability.

**Technology Acceptance.** Two main constructs, perceived ease of use and perceived usefulness, were selected from the Technology Acceptance Model (TAM; Davis, 1989) to measure students’ acceptance of the email digest as an awareness tool. A total of 12 items, including six items per construct, were modified to fit the research context of this study. The participants were asked to answer these items based upon their use of the email digest. The Cronbach $\alpha$ reliability estimates from our data were .96 for perceived ease of use and .98 for perceived usefulness.

**Self-Reported Participation.** The OLESQ includes four self-report items that ask students to estimate their participation through the use of tools in the course management system, such as discussion board, chat, resource sharing space, etc. The Cronbach $\alpha$ reliability estimates from our data were .73.

**Satisfaction.** Four learning satisfaction questions were taken directly from the Zone Experience Study Questionnaire (ZESQ; Lin, 2005). One question of learning interest and four course evaluation questions were adapted from ZESQ to be the five course evaluation items in OLESQ. These nine OLESQ questions measured students’ learning satisfaction and satisfaction with course materials and teaching in the online learning environments. The Cronbach $\alpha$ reliability estimates from our data were .87 for learning satisfaction and .89 for course evaluation.

**Data analysis**

The research questions were addressed through path analysis which helped us explore and determine the predictive relationships among social constructs. Since path analysis is a form of regression modeling for observed variables; therefore it requires normal regression assumptions, including univariate and multivariate outliers, linearity, normality, no multicollinearity, and homoscedasticity (Kline, 2005). Prior to path analysis, the data were examined for fit with the assumptions and no violations were found. Because the specification of directionality of path analysis requires a clear rationale and theory support, the literature review and correlation analysis were utilized to identify relationships among social constructs. According to the literature review, social constructs, such as sense of community, social ability, perceived ease of use, perceived usefulness, self-reported participation, and satisfaction, have interdependent relationships. The literature review provided insights about possible explanatory relationships among social constructs. Correlation analysis was implemented to identify significant relationships among social factors that could then be addressed in the path analysis model. Based on the literature review and correlation analysis, an initial path model of direct and indirect relationships among critical constructs was hypothesized and
examined. When exploring the predictive relationships, four correlated paths were discarded because their path coefficients indicated insignificant correlation (p<0.05). After examining the path models via Mplus®, a final model with best model fit was found to demonstrate the predictive relationships among social constructs of online learning.

Results

Prior to analysis, 1 of 52 cases was found to be a univariate outlier with an extremely low t score (t>3.29, α = .001) while no multivariate outliers were found (χ² (6) = 22.46, α = .001). The univariate outlier was deleted and left 51 cases for analysis. Also, the linearity of the variables was examined by checking scatter plots between any two of the 6 social constructs. A review of scatter plots indicated that every pair of variables had linear relationships. Further, the skewness and kurtosis of 6 variables were examined for normality. The skewness was found to be satisfactory (between .02 and -.47, < ±3) as well as the kurtosis (between 3.3 and -1.18, < ±3). Kolmogorov-Smirnov (K-S) values were found to be non-significant (greater than .01) and normal probability plot suggests no major deviations from normality. Overall, the data is distributed normally. Thus, there is no necessity for deletion or transformations of the variables. Additionally, Tolerance values with a cutoff value of 0.1 were utilized to assess multicollinearity. All the tolerance values were greater than 0.1 (between .39 and .65) which shows no violation of multicollinearity assumption. Lastly, a standardized residual plot indicated that the residuals were roughly rectangularly and evenly distributed around the 0 point of the standardized predicted value (X axis) which suggests no violation of the homoscedasticity assumption. These results show that the 51 cases meet the assumptions for path analysis.

Descriptive statistics for the research constructs are presented in Table 2. The findings indicate that students had positive perceptions of sense of community, social ability, perceived ease of use, perceived usefulness, self-reported participation, and satisfaction. Additionally, all constructs satisfied the criteria for reliability. Most constructs had Cronbach’s alpha values close to or over .80 (Nunnaly, 1978), while “self-reported participation” was the only construct showing only a moderate level of reliability (α = .73).

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Overall</th>
<th>Reliability (# of items)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of Community (SOC)</td>
<td>4.75</td>
<td>.94</td>
</tr>
<tr>
<td>Social Ability (SA)</td>
<td>5.21</td>
<td>.76</td>
</tr>
<tr>
<td>Perceived ease of use (PEU)</td>
<td>4.49</td>
<td>1.28</td>
</tr>
<tr>
<td>Perceived usefulness (PU)</td>
<td>4.09</td>
<td>1.45</td>
</tr>
<tr>
<td>Self-reported Participation (SRP)</td>
<td>5.00</td>
<td>1.11</td>
</tr>
<tr>
<td>Satisfaction (S)</td>
<td>5.49</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Table 3 presents a correlation matrix of all critical constructs. Students’ sense of community, social ability, self-reported participation, and satisfaction have highly significant positive intercorrelations. Students’ perceived ease of use with the awareness tool has significant positive correlation with sense of community, social ability, and satisfaction, while students’ perceived usefulness of the awareness tool only has significantly positive correlations with sense of community and satisfaction. Self-reported participation does not have significant correlations with students’ perceived ease of use or usefulness. To summarize, the results show that students’ technology acceptance (perceptions of use of the awareness tool) is associated with sense of community, social ability, and satisfaction but has no significant relationship with the self-reported participation through using tools in the Sakai system.

<table>
<thead>
<tr>
<th>Variables</th>
<th>SOC</th>
<th>SA</th>
<th>PEU</th>
<th>PU</th>
<th>SRP</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of community (SOC)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social ability (SA)</td>
<td>.609**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use (PEU)</td>
<td>.358*</td>
<td>.324*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness (PU)</td>
<td>.340*</td>
<td>.011</td>
<td>.570**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Reported Participation (SRP)</td>
<td>.516**</td>
<td>.313*</td>
<td>.069</td>
<td>-.040</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Satisfaction (S)</td>
<td>.786**</td>
<td>.573**</td>
<td>.316*</td>
<td>.303*</td>
<td>.563**</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. ** P<.01, *P<.05
The research path model was analyzed by Mplus@. In path analysis, the correlated paths of a prior path model are constructed based upon the significance of correlation coefficients. In the prior path model, social ability, perceived ease of use, perceived usefulness, and self-reported participation were hypothesized to predict students’ sense of community and overall satisfaction with online learning. Similarly, perceived ease of use, perceived usefulness, and self-reported participation were hypothesized to predict social ability. When the correlations from the prior path model were examined, four direct correlation paths (PEU → SOC, PEU → S, PU → S, and SA → LS) were discarded because of insignificant correlation coefficients (P<.05). When dropping these 4 paths, chi-square change was not greater than $\chi^2(0)=3.84$. After discarding the four non-significant paths, a final path model with best model-fit was found. The final model with R² values is presented in Figure 1. According to the criteria recommended by Hu and Bentler (1998), the chi-square value ($\chi^2$) for the model was 7.89 (p>.05) suggesting that the data fit the model well. The comparative fit index (CFI) was .99 and the Tucker-Lewis Index (TLI) was .97, which also suggested that the data fit the model well. However, the root mean square error of approximation (RMSEA) was .08 and the confidence interval of RMSEA is from .00 to .21 (include .05), which suggest a marginal fit of the model. The fit indices of goodness are presented in Table 4. Overall, the data fits the final model well.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>P</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>RMSEA 90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>N/A</td>
<td>&gt;.05</td>
<td>.90</td>
<td>.90</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Results of the Final Model</td>
<td>7.89</td>
<td>.25</td>
<td>.99</td>
<td>.97</td>
<td>.08</td>
<td>.00 ~ .21</td>
</tr>
</tbody>
</table>

In the final path model, students’ overall satisfaction with learning in online courses is explained by students’ sense of community, social ability, and self-reported participation. Additionally, social ability, perceived usefulness of social awareness tools, and self-reported participation were predictors of sense of community, which in turn predicted satisfaction. Moreover, perceived social ability is predicted by perceived ease of use and self-reported participation. Interestingly, students’ perceived ease of use of social awareness tools is a direct predictor of social ability and perceived usefulness but not sense of community, while perceived usefulness of social awareness tools is a direct predictor of sense of community but not social ability. Except for the lower correlation coefficient between self-reported participation and satisfaction (significant at p<0.05), the correlation coefficients range from .34 to .97 and are statistically significant at p<0.001. The R² s means indicated that approximately 66% of the variance in satisfaction is accounted for by sense of community and self-reported participation, 63% of the variance of sense of community is accounted for by social ability, perceived usefulness, and self-reported participation, 22% of the variance of social ability is accounted for by perceived ease of use and self-reported participation, and 33% of the variance of perceived usefulness of social awareness tools is accounted for by perceived ease of use.

Figure 1. Final Path Analysis Model with R² Values (* z≥1.96, p<0.05; ** z≥3.29, p<0.001 statistically significant; → represents significant path, ← represents variance explained)
Baron and Kenny (1986) defined mediator as a variable accounting for the relation between a predictor and a dependent variable. According to the steps for examining mediating relationships described in Cokley (2003), a mediator is identified when the relationship between a predictor and a dependent variable becomes insignificant while adding the mediator into the path relationship. Based on Cokely’s steps, three potential mediating relationships were examined. The results of identifying mediators in the present study show that sense of community mediated the relationships between social ability and satisfaction (SA → S) and between perceived usefulness and satisfaction (PU → S); however, sense of community did not mediate the relationship between self-reported participation and satisfaction (SRP → S). Figure 2 shows how sense of community serves as a mediator for in two relationships but not the third. When the model only includes paths from social ability to satisfaction (SA → S) and social ability to sense of community (SA → SOC), there are significant relationships between social ability and satisfaction (SA → S, b=.57**) and social ability and sense of community (SA → SOC, b=.61**). After adding the path from sense of community to satisfaction (SOC → S, b=.76**), the relationship between social ability and satisfaction (SA → S, from b=.57** to .15) decreased and became insignificant. Similarly, for the perceived usefulness model, the relationship between perceived usefulness and satisfaction (PU → S, from b=.30* to .04) decreased and became insignificant when the path between sense of community and satisfaction (SOC → S) was added. However, for the self-reported participation model, the relationship between self-reported participation and satisfaction (SRP → S, from b=.56** to .21*) decreased but remained significant when the path between sense of community and satisfaction (SOC → S) was added. Thus, there was no mediating relationship found for the self-report participation model.
Discussion

According to the social theory of learning (Wenger, 1998), members learn by participating in activities and socially interacting with other members of their community. Previous studies have found that students’ sense of belonging, communication styles, ways of appropriating tools, and having relationships with others influence their intentions toward participating in course activities and interacting with others. To better understand how social factors impact students’ learning interdependently in online learning environments, one research question of this study addressed the influence of sense of community on satisfaction with online learning experience. Our study followed the approach to measuring satisfaction found in other studies of distance education and online courses (Lin, 2005; Alavi, Wheeler, & Valacich, 1995). We assessed students’ satisfaction by measuring students’ course evaluations and perceptions of learning satisfaction. Consistent with previous studies (e.g. Dawson, 2006; Rovai, 2002a; Rovai, 2002c), our results indicated that, students’ perceived sense of community had a strong and positive influence on their overall satisfaction with online learning (Rovai, 2002a). In one study, Rovai (2002c) found that the online learners who had a stronger sense of community and perceived higher cognitive learning felt less isolated and had greater satisfaction with their academic program. Moreover, Dawson (2006) found that students who have more interaction with their peers as well as their instructor had higher degrees of sense community, and it influences them to have higher levels of satisfaction with their courses. It is critical to establish students’ sense of community in order to promote their participation and interaction in online learning.

In addition to previous studies, we examine the relationships among social constructs simultaneously. This examination shows how sense of community serves as a mediator for other social constructs (e.g. social ability and perceived usefulness of social awareness tools) to online learning satisfaction. The mediating role of variables shows the importance of examining social constructs simultaneously. Compared to previous studies which have examined one-by-one relationships between social constructs within online learning, we found the relationships among social constructs changed when examining the social constructs simultaneously. A path analysis indicated that sense of community mediates the influence of social ability and perceived usefulness of social awareness tools on satisfaction with online learning. In contrast with a previous study (Lin et al., 2006), we identified social ability as an indirect predictor of satisfaction with online learning, and determined that sense of community is a more direct predictor of satisfaction than social ability. Lin et al. (2006) examined the relationship between social ability and satisfaction without considering the interaction effects from other social constructs; however, we found the social constructs to be interdependent and intertwined. Rather than examining the relationship between social constructs separately, new insights can be gained by studying the intertwined relationships simultaneously.

Further, the final path model not only shows what social constructs contribute to building a sense of community and social ability but also reveals the importance of how students’ perceptions of social awareness tools are important to social ability and sense of community. Students’ acceptance of social awareness tools as easy to use contributes to their social ability and to their perceived usefulness of these tools. Aligned with Hara and Kling’s (2000) finding that students’ interaction in online learning was diminished when experiencing difficulties using tools, we found students’ social ability affected by their perceptions and feelings about the social awareness tools. Students need to feel that the tool is easy to use before they judge whether the tool is useful for their learning. Also, perceived ease of use was found to directly and positively influence students’ social ability, while students’ perception of tool usefulness played a direct role in their sense of community. These findings indicate that students appreciate the social awareness tool as they discover it is beneficial to their learning and increases connectivity among community members. However, although students can use tools easily to interact with others in class, if there is no sense of community then their participation and interaction tend to be restrained. This finding confirms the theoretical insights of the social theory of learning in that students’ growth not only includes cognition changes about a subject matter but also their identity and sense of belonging, which in turn influences students’ further participation and interaction. These findings also serve to remind system designers that a system needs to be developed beyond simple ease of use and usefulness for access to information to also consider supporting socio-emotional interactions that may promote a sense of community.

Finally, our findings illustrate the positive, direct impact of participation on both social ability and sense of community. This result suggests that the more students use the communication tools of the online learning environment, the more they may develop social ability and sense of community in the online context. Therefore, so as to increase social ability and sense of community, online instructors may want to thoughtfully encourage greater use of communication tools in their instructional designs. Although our findings extend previous literature on online
learning by revealing complex relationships among key constructs related to the social nature of online learning, these findings should be applied with care, as more research is required. We have three specific cautions. First, the sample used in this study is a convenience sample and is relatively small. The small sample size constrains the path analysis to include only 6 social constructs in the model, which left 5 sub-constructs of social ability out of the model. However, it would be interesting to investigate how sub-constructs of social ability related to other social constructs individually. Thus, having randomly drawn larger sample sizes would allow the addition of more parameters in path analysis and a more complete examination of the interdependent relationships among social constructs in online learning. Second, only self-reported data were used to measure key constructs used in the model. For example, self-reported participation was used to estimate students’ participation and it might not address the meaning of participation addressed by Wenger thoroughly. Self-reported participation can be different from actual participation and the participation perceived by others in classes. Further study should strive to include other data forms such as actual logs of participation activity and content analysis of communication and student products to measure students’ level of participation and interaction. Third, there are still some personal variables (e.g., age, gender, and academic level) and course variables (task types, course structures, and instructor experience) not addressed by this study. Further study should try to explore more about how different age, gender, or course related variables influence the relationships of social constructs regarding students’ online learning experiences. Addressing these limitations should increase the generalization of the findings about relationships among social constructs to other learning situations.

Educational Importance of the Study

Our study contributes to online education research in several ways. First, our findings support literature that shows sense of community is highly and positively related to student online learning satisfaction. Second, we have built new knowledge that advances the theory of how sense of community develops in online environments, which in turn provides guidance to foster and sustain students’ sense of community. The findings illustrate that sense of community is built through students’ participation in online activities, positive perceptions toward others in the course, and positive attitudes toward social awareness tools. These findings provide online instructors with an improved sense of how to best support students. Additionally, our findings also identified the impact and need for students’ perceptions of social awareness supports/tools influencing the development of students’ online learning satisfaction. The result showing that students’ perceptions of ease of use and usefulness of social awareness tools affect social ability and sense of community differently provide guidance for systems developers to improve mechanisms of awareness information. Lastly, the results of this study show the importance of examining the relationships among social constructs of online learning simultaneously. Path analytic techniques help uncover the interdependencies among constructs within an overall explanation of outcomes. Future research will consider how the sub-factors (e.g. social presence, social navigation, etc) associated with social ability might relate to other constructs in the model and how behavioral measures of participation can enhance the model. Additionally qualitative methods will contribute to broadening understanding about potential other factors as well as deepening understanding of what it means to be in an online learning environment under different social conditions.

Acknowledgements

We wish to thank the students who participated in this study and the SCRG (Social Computing Research Group) researchers at The University of Missouri, who helped with data collection. We would also thank Chris Amelung who developed the CANS (Context Awareness and Notification System) for students to use.

References


Appendix A: An example of a daily email digest

CANS notification digest for Friday, April 28, 2006

Intro Web Dev FS06
https://sakai.school.edu/portal/site/29c7cde6-19d6-4f75-0008-05e680708825

Discussion
Gary Arndt posted "Strength_Arndt" 16:00:18
Amy Burns posted "Re:For the Final Project" 13:43:40
Anna Duncan posted "Re:For the Final Project" 12:38:36
John Turner posted "Strengths_Duncan" 10:08:28

Resource
Amy Burns created "unit10_assignment.pdf" 11:09:29
Amy Burns created "unit10_assignment.pdf" 11:08:39
Amy Burns created "62a20e58-f083-4b47-0057-0884600251c8" 11:08:39

Chat
Amy Burns created 13:45:31
Allison Kim created 12:33:16
Amy Burns created 11:59:30
Amy Burns created 11:52:24
Amy Burns created 11:52:10
Allison Kim created 11:51:44
Allison Kim created 11:48:12
Amy Burns created 10:37:43

Announcement
Kimberly White posted 00:57:44
Amy Burns posted 00:40:32

This automatic notification message was sent to users of the CANS Email Digest (http://cans.school.edu) To unsubscribe, send an email to zone@school.edu
Appendix B: Online Learning Experience Study Questionnaire (OLESQ)

Directions: The following questions use the rating scales from “strongly disagree” to “strongly agree” as end points. Remember there is no right or wrong answer just answer as accurately as possible. For example, if you strongly agree with the statement, select 7; if you strongly disagree, select 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes your opinion.

Sense of Community
1. I feel that students in this course care about each other
2. I feel that I am encouraged to ask questions
3. I feel connected to others in this course
4. I feel that it is hard to get help when I have a question
5. I do not feel a spirit of community
6. I feel that I receive timely feedback
7. I feel that this course is like a family
8. I feel uneasy exposing gaps in my understanding
9. I feel isolated in this course
10. I feel reluctant to speak openly
11. I trust others in this course
12. I feel that this course results in only modest learning
13. I feel that I can rely on others in this course
14. I feel that other students do not help me learn
15. I feel that members of this course depend on me
16. I feel that I am given ample opportunities to learn
17. I feel uncertain about others in this course
18. I feel that my educational needs are not being met
19. I feel confident that others will support me
20. I feel that this course does not promote a desire to learn

Social Ability

Factor 1: Perceived peers social presence
1. I feel connected to other students in this course
2. My interactions with other students are sociable and friendly
3. My online interactions with other students seem personal
4. In my interactions with other students I am able to be myself and show what kind of classmate I really am
5. I feel like I am a member of a group during the course activities
6. I feel comfortable expressing my feelings to other students
7. When I log on I am usually interested in seeing what other students are doing or have done
8. I trust the other students in this course to help me if I need it
9. The actions of other students in the course are easily visible in our online system
10. When I see that other students are confused I offer help

Factor 2: Perceived written communication skills
11. I am concerned that my writing ability limits how well other students can get to know me
12. I am concerned that my writing ability limits how well my instructor can get to know me
13. I am concerned that my writing ability limits how effective I can be in this course

Factor 3: Perceived instructor social presence
14. My interactions with the instructor are sociable and friendly
15. I feel comfortable expressing my feelings to the instructor
16. My online interactions with the instructor seem personal
17. The actions of the instructor in the course are easily visible in our online system
18. In my interactions with the instructor I am able to be myself and show what kind of student I am really am
19. I trust the instructor in this course to help me if I need it
20. When I log on I am usually interested in seeing what the instructor is doing or has done
21. I feel connected to the instructor in this course

**Factor 4: Comfort with sharing personal information**
22. I feel uncomfortable with the amount of information about myself that I had to share with other students in this course
23. I feel uncomfortable with the amount of information about myself that I had to share with the instructor in this course
24. I feel uncomfortable interacting with others in the course because these interactions are recorded

**Factor 5: Social navigation**
25. Knowing what other students in the course have done helps me know what to do
26. Knowing that other students in the course are aware of my work usually influences how hard I work and the quality of my work
27. The actions of other students in the course influence the quality of my work (such as trying to write better messages or working more carefully)
28. Interacting with the instructor helps me accomplish assignments with higher quality than if I were working alone
29. Interacting with other students helps me accomplish assignments with higher quality than if I were working alone
30. The actions of the instructor in the course influences the quality of my work (such as trying to write better messages or working more carefully)

**Technology Acceptance: Perceived Ease of Use**
1. Learning to use Notification (widget or digest) was easy for me
2. Notification was flexible to interact with
3. I find it's easy to get Notification to do what I want to do
4. It's easy for me to become skillful at using Notification
5. My interaction with Notification is clear and understandable
6. Notification is easy to use

**Technology Acceptance: Perceived Usefulness**
1. Using Notification helps me learn about & accomplish the course requirements quickly
2. Using Notification helps me to be a productive student
3. Using Notification enhances my effectiveness in the course
4. Using Notification makes it easy to do the course coursework
5. Using Notification improves my academic performance
6. I find Notification a useful tool for my learning in this course

**Directions:** The following questions use the rating scales from “never used” to “a great extent” as end points. Throughout this questionnaire we use the term “courseware.” By courseware we mean the software system you use for your course, e.g., Sakai, Blackboard or WebCT. Remember there is no right or wrong answer just answer as accurately as possible. For example, if you have performed the action a lot, select 7; if you have never performed the action, select 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes your level of activity.

**Self-Reported Participation**
Please indicate the extent to which you did the following Course-Related activities
1. Posted messages on the discussion board
2. Read messages on the discussion board
3. Sent Emails to others
4. Uploaded files into the courseware system

**Directions:** The following questions use the rating scales from “strongly disagree” to “strongly agree” as end points. Remember there is no right or wrong answer just answer as accurately as possible. For example, if you strongly
agree with the statement, select 7; if you strongly disagree, select 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes your opinion.

**Satisfaction with Learning in Online**
1. I developed knowledge and competencies in this course
2. The course activities were a good fit for the way I like to learn
3. The course activities met my expectations for what I had hoped to learn
4. The knowledge and competencies taught through the course activities are personally meaningful and important to me
5. Course learning objectives were clear
6. I usually have a clear idea of where I am going and what is expected of me in this course
7. The teaching materials for this course are good at explaining things
8. The course tries to get the best out of all the students
9. I’ll be happy to take another online course similar to the one I am taking now
Considering Students’ Perceptions: The Distance Education Student Satisfaction Model

Ismail Sahin
Department of Computer and Instructional Technologies, College of Education, Selcuk University, Turkey // isahin@selcuk.edu.tr

Mack Shelley
Iowa State University, USA // mshelley@iastate.edu

ABSTRACT
In the current study, the Distance Education Student Satisfaction Model, estimated as a structural equation model, is proposed to understand better what predicts student satisfaction from online learning environments. In the present study, the following variables are employed based on the Technology Acceptance Model (TAM) (Davis, Bagozzi, & Warshaw, 1989) and literature: computer knowledge, flexibility of distance education, usefulness of distance education, and distance education satisfaction. Results suggest that as long as students have the skills to use online tools and perceive that distance education is a useful and flexible way of learning, communicating, and sharing, their enjoyment from online instruction will be promoted. Ultimately, this satisfaction may lead to higher levels of engagement, learning, and success in the distance education setting. Data collected from 195 undergraduate students are analyzed using Statistical Package for the Social Sciences and Analysis of Moment Structures statistical software. Implications of the findings of the present study are crucial for instructors, practitioners, and institutions planning to offer or currently engaged in offering distance education courses.

Keywords
Distance education, Student satisfaction, College of education

Schools at all levels, and institutions of higher education, have invested large budgetary and other resources in online learning tools, especially the Internet (Cheung & Huang, 2005; Rafaeli et al., 2004) since it has become a valuable tool for teaching and learning worldwide and “has brought dramatic changes to education in general and distance learning in particular” (Holsapple & Lee-post, 2006, p. 68). Now, the Internet is incorporated into educational settings to extend learning activities without depending on traditional classroom space and time (Hagel & Shaw, 2006; Jones et al., 2004; Xie, Debacker, & Ferguson, 2006). In fact, flexibility of time and place for learning is the most important feature of online instruction. While distance education provides an interactive, reflective, and collaborative learning setting (Maor, 2003), it also challenges researchers and designers of online instruction to develop educational software and find ways to support online learning environments (Ardito et al., 2006; Green, 2006; Yang & Cornelious, 2005) in which students’ needs are fulfilled.

There are many ways in which Web tools can be used to support teaching and learning (Zhang, Perris, & Yeung, 2005; Frederickson, Reed, & Clifford, 2005). The contributions of distance education to individual learners may include (White, 2005):
- broadening access to education
- providing new learning environments
- individual development
- knowledge and awareness of learners in context
- the importance of understanding the learner’s perspective on distance education

It is clear that online education has the potential to provide students with high-quality learning experiences. If the course content is prepared by considering students’ value system, along with their social and cultural context, learning is more likely to occur (Bradshaw & Hinton, 2004; Levin & Wadmany, 2006; Muilenburg & Berge, 2005). The literature criticizes the assumption that most students have the ability to use the information and communication technologies within an educational setting (Jones et al., 2004) and suggests that many undergraduate students entering the university have no experience with the Internet and very little with information technology generally (Arif, 2001). In fact, distance education tools might seem to be unfamiliar or difficult to learn for many students, so they might not be enthusiastic about participating in online activities (Hong, Ridzuan, & Kuek, 2003; Xie et al.,...
Hence, it is essential that students should have basic computer skills to maintain control of their own learning in distance education.

Much research has been conducted on distance education (Chambers, 2006; Hagel & Shaw, 2006; Hong et al., 2003; Lee et al., 2003; Liao, 2006; Muilenburg & Berge, 2005). The literature emphasizes the importance of research into student learning for professional practices of course designers and tutors, and for improving students’ distance learning experiences (Levin & Wadmany, 2006; White, 2005). In designing, developing, and delivering distance education courses, students’ needs and perceptions should be central. A course failing to meet student expectations and needs may lead to low levels of student involvement (Hall, 2001). Indeed, without investigating what satisfies undergraduate students in distance education courses, it is difficult to meet their needs and improve their learning. We conceptualize satisfaction from technology as an essential link in student outcomes, with greater enjoyment associated with higher levels of student engagement, which in turn is associated with increased student learning.

Moreover, distance education involves a student-centered approach in which the instructor takes the role of the facilitator and students engage in peer learning (Maor, 2003; Mitchell, Chen, & Macredie, 2005). It is clear that Web-based learning environments having a student-centered approach should consider student satisfaction. Overall, the literature suggests that there is a need to understand better the variables that affect student enjoyment of distance education courses. This article reports the results of a research study investigating predictors of student satisfaction in online learning environments. In this research, the Technology Acceptance Model (TAM) informed the selection and measurement of variables. As shown below, students’ satisfaction is operationalized here in terms of their assessment of student-centered instruction, class content, real-life relevance, advising others to take the class, and timeliness of responses to questions about class content (Cronbach’s standardized item alpha = 0.78). Relationships between satisfaction and other derived factors are examined in our statistical analysis.

One of the most-cited models (Lee et al., 2003), the TAM is proposed to understand the user acceptance of information systems (Davis, Bagozzi, & Warshaw, 1989). It provides a basis to explain the impact of variables such as beliefs, attitudes, and intentions on using a technological application. In the TAM, computer use is determined by behavioral intention that is formed by perceived usefulness and attitude. Perceived usefulness is described as “the prospective user’s subjective probability that using a specific application system will increase his or her job performance within an organizational context” (Davis et al., 1989, p. 985). The traditional TAM framework measures perceived usefulness and perceived ease of use. Here, we adopt and apply Davis et al.’s definition of perceived usefulness, and investigate the role of satisfaction with information technology, as suggested in previous research.

Internet use might be stimulated by two dimensions of motivators: extrinsic (perceived usefulness) and intrinsic (satisfaction) (Cheung & Huang, 2005). In fact, the literature shows that both perceived usefulness and perceived satisfaction directly affect undergraduate students’ intention to use Internet-based learning resources (Lee, Cheung, & Chen, 2005). Two additional motivators are proposed by Straub (1994): social presence and information richness. In the present study, which evaluates student perceptions of distance education, social presence is represented by the flexibility of distance education and information richness is represented by computer knowledge. In the current study, the TAM and related literature are used to construct the following variables: computer knowledge, flexibility of distance education, usefulness of distance education, and distance education satisfaction. The relationships among these variables were examined in a structural equation model.

Methodology

The data were collected from undergraduate students of the College of Education at an Anatolian university in Turkey. Distance education tools, such as e-mail messages, discussion boards, online assignment submissions, and online exams, were used to support their learning in the class. The course, Introduction to Educational Technology, was offered during the 2005-2006 academic year.

Participants

The participants of this study consisted of 195 undergraduates. Of the participants, 60% were male (n = 118) and 40% female (n = 77). The average age for the participants was 21 years. Almost half of the students (46%) reported that they own a computer.
Research Instrument

This study analyzed self-reported opinions and perceptions of Turkish undergraduate students regarding distance education measured by responses to an online survey. The survey contained three selected participant demographics: gender, age, and computer ownership. In addition to participant demographics, the survey included Likert-type items with response choices ranging from “strongly disagree” to “strongly agree.”

Computer expertise

The first section consisted of items about the role of the online course in increasing the students’ computer skills and use of the Internet. Higher scores in this part indicated higher perceived computer knowledge and use.

Flexibility of distance education

This subscale assessed students’ perceptions of the flexibility characteristic of distance education. Higher scores reflected more positive beliefs about the adaptability of distance education.

Usefulness of distance education

This section included items regarding the participants’ attitudes toward the usefulness of distance education. Higher scores indicated more strongly positive beliefs about the value of distance education.

Distance education satisfaction

This subscale measured how much the students were satisfied with distance education. Higher scores showed higher levels of satisfaction from distance education.

Data Analysis

Exploratory factor analysis was used to verify whether the survey items for each subscale successfully measured each variable. In addition, reliability analysis, which assesses the internal consistency among sets of survey items (Mertler & Vannatta, 2002), was employed to measure the reliability of each section of the survey. Reliability was measured using standardized Cronbach’s alpha values, for which a level of .8 or above often is taken to denote strong split-half consistency. Next, structural equation modeling (SEM) procedures were used to explore relationships among the variables. For each endogenous (dependent) variable, an equation was estimated by exogenous (independent) or other endogenous variables from another equation. Both the direct and indirect effects of independent variables on the dependent variables were estimated. The structural model was tested by examining the path coefficients, which are standardized regression coefficients (betas). Statistical analyses were conducted using Statistical Package for Social Sciences and Analysis of Moment Structures (AMOS) software.

Findings

Exploratory Factor Analysis (EFA)

In the EFA, the component matrix showed positive, high correlations among the separate items included within each of these four factors (see Appendix A for the results of factor loadings and reliability tests). The KMO statistic and Bartlett's test results ($p < .001$) showed that the items constituting each of the four factors fit together appropriately for the results of EFA to be meaningful. Also, the value of the Cronbach standardized item alpha for each factor was high. These results verified that a summated rating scale for each factor can be formed meaningfully from each of these sets of variables.
Items with loadings equal to or greater than 0.50 were retained on each factor. The first factor, which included four items with reliability of 0.86, was named “computer knowledge.” The second factor (“flexibility”) consisted of four items with reliability of 0.66. The third factor, named “usefulness,” was composed of five variables with reliability 0.83. The last section of the survey, dealing with student pleasure with distance education, included five items and its Cronbach standardized item alpha value was 0.78. This variable was labeled “satisfaction.”

**Structural Equation Analysis: The Distance Education Student Satisfaction Model**

The structural equation analysis was conducted to test the relationships among computer knowledge, flexibility of distance education, usefulness of distance education, and distance education satisfaction. As depicted in Figure 1, the full AMOS model, which fits the data perfectly, included one exogenous variable (computer knowledge) and three endogenous variables (flexibility, usefulness, and satisfaction). Although these variables were generated through EFA, and hence represent latent traits, they are denoted in Figure 1 as rectangles because their values were calculated as the means of the survey items that loaded most strongly on each factor and each is summarized by a single composite mean score.

![Figure 1. Distance Education Student Satisfaction Model](image)

As seen in Table 1, total effects were decomposed into direct and indirect effects. In the model, all path coefficients were statistically significant.

**Table 1. Decomposition of Total Effects for Model**

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Dependent Variable</th>
<th>Total Effect</th>
<th>Direct Effect</th>
<th>Indirect Effect</th>
<th>Standard Error</th>
<th>Critical Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Expertise</td>
<td>Flexibility of Distance Education</td>
<td>0.51</td>
<td>0.51</td>
<td>0.00</td>
<td>0.06</td>
<td>8.18**</td>
</tr>
<tr>
<td>Computer Expertise</td>
<td>Usefulness of Distance Education</td>
<td>0.50</td>
<td>0.21</td>
<td>0.29</td>
<td>0.06</td>
<td>3.54**</td>
</tr>
<tr>
<td>Flexibility of Distance Education</td>
<td>Usefulness of Distance Education</td>
<td>0.57</td>
<td>0.57</td>
<td>0.00</td>
<td>0.06</td>
<td>9.64**</td>
</tr>
<tr>
<td>Computer Expertise</td>
<td>Distance Education Satisfaction</td>
<td>0.61</td>
<td>0.32</td>
<td>0.29</td>
<td>0.06</td>
<td>5.65**</td>
</tr>
<tr>
<td>Flexibility of Distance Education</td>
<td>Distance Education Satisfaction</td>
<td>0.48</td>
<td>0.35</td>
<td>0.13</td>
<td>0.07</td>
<td>5.28**</td>
</tr>
<tr>
<td>Usefulness of Distance Education</td>
<td>Distance Education Satisfaction</td>
<td>0.22</td>
<td>0.22</td>
<td>0.00</td>
<td>0.07</td>
<td>3.36**</td>
</tr>
</tbody>
</table>

*: Total effect = Direct effect + Indirect effect; **: p < 0.01
The positive direct effect between computer expertise and flexibility was statistically significant ($r = .51, p < 0.01$). Students who perceive that they possess a higher level of computer knowledge ($t = 8.18, p < 0.01$) stated that they have more positive beliefs about the flexibility of distance education; $26\%$ of the variation in flexibility was explained by computer expertise ($p < 0.01$). Thus, knowledge of students’ computer skills and use significantly increases the ability to predict students’ attitudes toward the adaptability of distance education. Of course, there may be a difference between students’ perceptions and reality; perception, by definition, is based on one’s own criteria and perspective. Some students who perceived that they had a higher level of computer expertise actually may have less computer knowledge than at least some of the students who perceived themselves as having a lower level of computer knowledge.

In the model, computer expertise affects the usefulness of distance education both directly and indirectly. Students who had more computer knowledge ($t = 3.54, p < 0.01$) and realized the flexibility of distance education ($t = 9.64, p < 0.01$) were more likely to find distance education useful. Computer expertise and flexibility explained $49\%$ of the variation in usefulness ($p < 0.01$).

Computer expertise and flexibility influence distance education satisfaction both directly and indirectly. Students who had more computer knowledge ($t = 5.65, p < 0.01$), recognized the flexibility of distance education ($t = 5.28, p < 0.01$), and described distance education as useful ($t = 3.36, p < 0.01$) were more likely to satisfy from distance education. The combination of computer knowledge, flexibility, and usefulness accounted for $57\%$ of the variation in distance education satisfaction ($p < 0.01$).

**Discussion**

The findings from the current study show the importance of computer knowledge and attitudinal factors in predicting student satisfaction from their distance education experience. The literature supports our finding that a higher level of computer experience is linked to greater enjoyment of users with Web-based learning (Mitchell et al., 2005). In fact, “inadequate or incomplete knowledge and awareness inevitably compromises the quality and appropriateness of learning experiences which can be provided and developed” (White, 2005, p. 170). Students’ online readiness, motivation, and attitude are keys for the success of any training program (Ardito et al., 2006; Holsapple & Lee-post, 2006; Xie et al., 2006; Zhang et al., 2005). Hence, technical issues should not become a barrier to students’ learning. It is crucial to note that a high level of computing experience—and not technical issues—helps students focus on learning. Students who use IT in their personal and professional lives are more comfortable and familiar with the online learning environment (Jones et al., 2004; Maor, 2003).

To fulfill students’ expectations from online learning environments, they need to be supported both technically and technologically. Institutions and educators should create opportunities and devote resources to assist students in developing their computer skills and expertise needed for online learning. Before offering a distance education course, the instructor should make sure that the students have basic computer skills so they will not be frustrated and discouraged with using the tools and environments of the online class. If necessary, at the beginning of the semester, the students who have a low level of computer proficiency should be provided with a training program to assure that they gain the computer skills required for the distance education course.

The results of this study highlight the critical role of perceived usefulness in student satisfaction from Web-based learning environments. It is essential that students should value computers and distance education. If students believe that the distance education course is useful, they will be more likely to enjoy that class. It is clear that students’ enjoyment of distance education is a critical factor in their acceptance and use of the online learning environment (Lee et al., 2005; Mitchell et al., 2005). The literature suggests that perceived usefulness and perceived satisfaction may increase students’ use of Internet resources (Lee et al., 2005; Mitchell et al., 2005). Thus, heightened student satisfaction from their distance education experience may lead to further engagement in class activities, and eventually in higher levels of use of distance learning environments.

The findings from this study clearly show that a distance education course should provide students with great flexibility in interacting with their instructor, classmates, and the course content. Indeed, flexible course structure is a key strategy to overcome the intrinsic and extrinsic barriers to e-learning (Jones et al., 2004). As long as students have the skills to use online tools and perceive that distance education is a useful and flexible way of learning,
communicating, and sharing, their enjoyment from online instruction will be promoted. Ultimately, this satisfaction might lead to higher levels of engagement, learning, and success in the setting.

Conclusions

Understanding students’ perceptions regarding distance education is the first step for developing and implementing a successful online learning environment. The primary contribution of this research is in furthering our understanding of the variables that affect student satisfaction with classroom technology. The model presented in this study is formulated in an attempt to explain student satisfaction by using a small number of variables suggested by the TAM (Davis et al., 1989) and literature. According to the TAM, the availability of the four motivators (computer knowledge, flexibility of distance education, usefulness of distance education, and distance education satisfaction) may lead to higher levels of use of distance education.

Findings and implications from the current study suggest that students’ computer knowledge and perceptions, such as the perceived usefulness and flexibility of distance education, should be considered as predictors of their satisfaction from classroom technology, and ultimately for their success in online learning environments. Instructors of distance education need to focus upon preparing students to use a variety of computer technologies and be aware of the benefits of online learning. Thus, there is a need for well-designed and carefully implemented online learning environments that meet the needs and expectations of students. Web-based learning environments can be facilitated through activities that increase students’ level of computer knowledge and emphasize the flexibility and usefulness characteristics of distance education. The results of this study suggest that these factors are necessary to promote student satisfaction with distance education. Hence, institutions and instructors should pay special attention to the variables affecting student enjoyment with distance education courses. Then, the learning environment might be more intuitive, engaging, and ultimately more didactic.

In addition, it should be noted that the model proposed in the present study may fit well and be defensible with the contextually-related studies and literature, but it does not imply causality. Further research is needed to confirm the validity of the model. Thus, future studies might focus on its applicability to other areas and settings. These efforts will be beneficial to the improvement of the model and the continued development of research and practice in distance education. There is no doubt that many factors may influence student satisfaction in online learning environments. Further refinement of the proposed model could identify more factors that educators can influence to improve the quality of their courses. Although more than half (57%) of the variance in student satisfaction is explained by the model, additional survey items or variables, such as environmental and institutional measures, might be included in the analysis to extend our understanding of student satisfaction with distance education.

References


## Appendix A: Results of Exploratory Factor Analysis and Reliability Analyses

<table>
<thead>
<tr>
<th>Variables</th>
<th>Items</th>
<th>Factor Loadings</th>
<th>KMO and Bartlett</th>
<th>Standardized Item Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computer Expertise</strong></td>
<td>This course helps me use the Internet sources more efficiently.</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>My use of computers increases after taking this class.</td>
<td>0.88</td>
<td>0.82***</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>This course contributes to my knowledge of searching on the Internet.</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>My computer knowledge increases with the course assignments and projects.</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flexibility of Distance Education</strong></td>
<td>Distance education allows me to allocate my time better.</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance education allows me to work at home comfortably.</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>In terms of use of time and location, distance education is flexible.</td>
<td>0.69</td>
<td>0.72***</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Distance education is appropriate to students with different learning capacities.</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Usefulness of Distance Education</strong></td>
<td>I believe distance education is useful.</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A degree in distance education is valuable like a degree in traditional education.</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance education provides me with a valuable learning experience.</td>
<td>0.76</td>
<td>0.80***</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Distance education minimizes the inequalities in education.</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evaluation of the success in distance education is quite objective.</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Distance Education Satisfaction</strong></td>
<td>The student-centered instruction offered in this class is enjoyable.</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The content of this class meets my expectations.</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I like its content from real life.</td>
<td>0.67</td>
<td>0.80***</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>I advise other students to take this class.</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I am pleased with the timely responses to my questions regarding the class content.</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***: $p < 0.001$
A study of the relationship between student social networks and sense of community

Shane Dawson
Centre for Learning Innovation, Queensland University of Technology, Australia // sp.dawson@qut.edu.au // Tel: +61 7 3138 5417 // Fax: +61 7 3138 3986

ABSTRACT

As the integration of community-centred teaching practices intensifies, an understanding of the types of relationships that manifest in this network and the associated impact on student learning is required. This paper explores the relationship between a student’s position in a classroom social network and their reported level of sense of community. Quantitative methods, such as Rovai’s (2002b) Classroom Community Scale and social network centrality measures, were incorporated to evaluate an individual’s level of sense of community and their position within the classroom social network. Qualitative methods such as discussion forum content analysis and student interviews were adopted to clarify and further inform this relationship. The results demonstrate that the centrality measures of closeness and degrees are positive predictors of an individual’s reported sense of community whereas, betweenness indicates a negative correlation. Qualitative analyses indicate that an individual’s pre-existing external social network influences the type of support and information exchanges an individual requires and therefore, the degree of sense of community ultimately experienced. The paper concludes by discussing future recommendations for teaching practices incorporating computer-mediated communications.

Keywords
Social networks, learning communities, computer-mediated communication

Introduction

Recent education literature has highlighted the importance for practitioners to adopt a community-centred pedagogy as a strategy for facilitating student learning (e.g. Brook & Oliver, 2003; Fink, 2003; Johnson, 2001). The dominant tenet of this pedagogical approach can be traced back to the works of Dewey (1938/1963) and Vygotsky (1978) who maintained that the process of learning is facilitated through individual participation in social interactions. This pedagogical model is framed within social-constructivist principles with a focus on developing activities that promote learner-to-learner interactions to support the co-construction of knowledge and the sharing of information and resources. In this context, learning activities involving group work and collaboration are commonly implemented practices.

However, opportunities for the contemporary learner to engage with peers in a collaborative environment are problematic given the spatial and temporal requirements associated with traditional classroom settings (Squire & Johnson, 2000). The integration of online technologies, such as computer mediated communication (CMC), within the education sector can be seen as one approach for addressing these challenges and therefore, facilitate the implementation of collaborative learning activities. For instance, the adoption of CMC software provides individuals with the capacity to interact via computer networks regardless of spatial and temporal limitations (Kreijns, Kirschner, & Jochems, 2002). De Wever, Schellens, Valke, and Van Keer (2006), suggested that an additional advantage underlying the integration of asynchronous CMC is the capacity for students to reflect on postings and access additional resources before (re)contributing to the overall discussion and therefore facilitating the development of higher order learning outcomes.

The benefits derived from implementing CMC also extend to faculty and researchers as a source of evaluative data. As Meyer (2004) has noted, written communication exchanges occurring among learners are readily accessible for future review. Ahern, Peck, and Laycock (1992) analysed CMC transcripts when investigating the impact of moderator intervention techniques on student participation. In examining the interactions in lieu of the specific written content, Garton, Haythornthwaite and Wellman (1997) have demonstrated that the communication exchanges conducted via CMC can also be used to form a representation of the social network and identify potential relational patterns. The interrogation of these relational networks may inform education practitioners of the extent of community experienced among the student cohort and the progress and outcome of implemented learning activities.
While education research has primarily focused on developing a greater understanding of the learning process and the activities that promote learning in an online environment (e.g. Gunawardena, 1995; Hara, Bonk, & Angeli, 2000; Schellens & Valcke, 2006; Vonderwell, 2003) there has been limited research examining the types of relations and networks that develop within the education milieu (Cho, Gay, Davidson, & Ingraffea, 2007) and the impact these networks have on an individual’s psychological sense of community. The present study seeks to contribute to this field of knowledge by investigating the relationship between a student’s position in the social network and their overall sense of community. To address this aim, the paper firstly discusses the application of social network analysis (SNA) as a methodology for education studies. The paper then reports on the findings of an initial study juxtaposing Rovai’s (2002b) sense of community scale with SNA in an education environment.

Social Network Analysis

The concept of social network analysis (SNA) has attracted much attention in the social and behavioural sciences. Wasserman and Faust (1994) attributed this interest to the ability of SNA to describe the relationships that occur among social actors and the associated patterns arising from these interactions. Social network analysts actively investigate the exchange of resources between social actors and how these interactions afford the establishment of relationships within a social system (Haythornthwaite, 2002). From these investigations, analysts map and visually display the interactions that occur in order to ascertain the emergence of two types of patterns – social groups and social positions. Social groups relate to the collection of actors and the associated social interactions that arise from participation in the social system and social positions refer to the classifications of the sets of actors that are similarly linked to the social system (Freeman, 2000).

SNA has also been applied in community studies to investigate the social relationships and patterns that evolve through member interactions (Cho, Lee, Stefánone, & Gay, 2005; Haythornthwaite, 2002; Reffay & Chanier, 2002). The application of SNA to community studies typically involves an investigation into the types of resources exchanged, frequency, quantity and flow in order to describe the elements comprising the social system (Haythornthwaite & Wellman, 1998). For example, Reffay and Chanier (2003) incorporated SNA to describe relationships and interactions occurring among students and staff participating in computer-supported collaborative learning. The authors maintained that the use of SNA provides an opportunity to gauge the communicative interactions that take place and to assess the degree of cohesion within the group. Girvan and Newman (2002) have drawn similar conclusions regarding the applicability of SNA to ascertain the degree and strength of social ties developed within a social system in developing models of community structures.

Social ties and social capital

Central to SNA studies is the notion of strong and weak social ties. Granovetter (1982) argued that the social bonds established between individual actors can be classified as either strong or weak ties. Strong ties are often represented in close knit networks such as personal friendships and relations. In contrast weak ties are seen as linkages between different networks. Granovetter (1982) described weak ties as bridges - linking separate cliques into a larger network. Furthering this research on weak and strong ties in social networks, Haythornthwaite and Wellman (1998) demonstrated that the strength of the social bonds formed between actors in a social network influences an individual’s choice of communication medium, and the frequency of interaction. Individuals with weak ties tend to adopt more socially rich mediums, such as face to face meetings, as the mode of communication and interact less frequently, in contrast to individuals exhibiting strong social bonds.

Social capital has gained wide acceptance as a theory for understanding the social rules and relations intertwined within the social structures of communities, and more broadly society (Lin, 2001). Putnam (1993) defined social capital as the “features of social organisation, such as networks, norms and trust that facilitate coordination and cooperation for mutual benefit” (p. 67). Lin (2001) proposed an analogous definition of the term noting that social capital is an: “investment in social relations with expected returns in the marketplace”(p. 19). While this definition implies an economic return on initial investments, the marketplace may represent a variety of contexts such as political, economical, or importantly for this study, educational.
In the education context, students potentially invest in developing social relations for returns of support, both personal and academic in order to progress their individual academic goals (Cho et al., 2007). Students bring to class a wide subset of pre-existing social relations that are external to the education environment. In this context the education setting forms a small component of a student’s social system. Consequently, students have a wide degree of potential social capital that can be drawn upon to assist in developing an understanding of the subject content.

The identification of the position a student occupies within the established social network may inform practitioners of the role that an individual plays in the co-construction of knowledge and the types of resources and support they require. By viewing online communities as comprised of a set of social networks (Wellman et al., 1996) SNA affords a method to examine the overt interactions that occur in the online environment, the strength of those interactions, and the types of resources exchanged that foster the development of communities online (Haythornthwaite, 2002). As the formation of a cohesive social network is integral for effective social learning, the examination of the formation and structure of these social networks and the sense of community experienced among the student cohort provides new approaches to informing practitioners of the effectiveness of implemented practices. The present study investigated the relationship between sense of community and an individual’s position within a social network in a higher education environment. Specifically the study examined the following research question:

Is the composition of social networks evolving from a unit discussion forum related to the sense of community experienced among the student cohort?

Methodology

Study overview

This research study forms a component of a larger investigation examining the relationship between student communication interactions and sense of community in the higher education environment. To answer the research questions the study incorporates a mixed method approach utilising both quantitative and qualitative measures. Quantitative methods, such as Rovai’s (2002b) Classroom Community Scale and SNA centrality measures, were incorporated to evaluate an individual’s level of sense of community and their position within the social network. Qualitative methods such as discussion forum content analysis and student interviews were adopted to clarify and further inform the relationship between an individual’s position in the social network and their perceived sense of community.

Education students enrolled in second semester (July- December) 2005 undergraduate and postgraduate courses at a large metropolitan university were invited to participate in the study. All teaching units (N = 25) selected for the study contained an online component of supplementary learning resources including CMC software (e.g. discussion forum, listserve, email, synchronous chat). The term unit is defined as a focussed course of study of approximately 13 weeks in duration comprising a part of a larger program leading to a specific degree.

The sampled teaching units were available for both external and internal study modalities. An external modality is defined as study undertaken off campus. In this instance more traditional teaching methodologies such as face to face lectures are excluded. Teaching content and associated learning activities are delivered and accessed via the institution’s online environment. An internal modality refers to study undertaken on campus. Traditional modes of education delivery such as face to face lectures, tutorials, and workshops were included, with additional online resources and learning activities also integrated into the subject curriculum. In this context, an internal study modality can be seen to be more analogous to a blended model of learning.

Methods of data collection

The quantitative methodology comprised the primary source of data collection for this study. Quantitative data were derived from both student self-report responses to an online survey and student online user-behaviour with the University’s in-house learning management system (OLT). Online user behaviour was generated automatically using the OLT evaluation system. The data generated by this system permits an insight into the way learners and designers interact with the online learning environment, such as who (student or teaching staff) is posting and replying to the discussion forum; or the time and frequency of interactions.
Qualitative methodologies such as a case study and content analyses of discussion forum postings were also undertaken in order to verify conclusions derived from the quantitative component. Yin (1993) suggested that in instances where the phenomenon under investigation is complex and interwoven with the context, the incorporation of multiple data sources provides a method of triangulation. Studies of community and social networks represent such phenomena. The diversity of data sources included in this study can be seen to be complementary and enable the verification of each unique data set and the subsequently derived conclusions (Silverman, 2000).

**Sense of community**

Data relating to sense of community were collected from student responses to an online survey that was administered via the teaching units’ online space and via email notification. The online survey was based on Rovai’s (2002b) Classroom Community Scale (CCS). The CCS consists of 20 self-report items designed to evaluate the degree of community an individual experiences within an education setting. The scale comprises two subscales termed social community and learning community. The subscale social community relates to the students’ perceived levels of belonging, trust and cohesion. The learning community subscale is defined as the degree to which students share similar learning values and goals.

An initial pilot study (N = 160) was undertaken to confirm the survey instrument’s factorial validity. The constructs emerging from the exploratory factor analysis were found to be equivalent to Rovai’s (2002b) study. Additionally, reliability and internal consistency analyses were conducted to ensure generalisability of results. Cronbach α and Guttman split-half for the instrument was 0.90 and 0.89 respectively, indicating excellent reliability and consistency. Analysis of the two subscales also demonstrated excellent reliability and consistency with a 0.86, 0.85 (Cronbach α, Guttman split half respectively) for social community and 0.84, 0.76 (Cronbach α, Guttman split half respectively) for learning community. As the survey instrument demonstrated factorial validity and excellent reliability, the survey was then offered to the broader study participants (N=464).

**Social Network Analysis**

Social network analyses were applied to the captured communication logs (discussion forum posts) in order to investigate the level of participation and relationships manifesting within the sampled units. Although other additional communicative software such as chat and listserv were incorporated into the teaching units, these types of CMC were not widely utilised by a diversity of students or were merely used by the teaching staff as a means for updating and informing the student cohort. In contrast, the discussion forums were widely adopted by both students and teaching staff in a conversational manner. From this “direct active communication” (Harrer, Zeini, & Pinkwart, 2005, p. 192) data social relationships were interpreted from the identification of individual nodes and ties within the network. Nodes are defined as the individual actors while ties refer to the bonds that link actors in a network (Freeman, 2000).

Unit discussion forum logs were extracted as xml files for further network analysis using NetVis 2D (NV2D), a java based graph visualisation tool. The open source software processes the discussion forum xml files and generates a sociogram based on the communication interactions occurring between students. The software has the capacity to calculate the various SNA centrality measures for each node generated through the initial graph visualisation step (Figure 1).

Social network calculations such as betweenness, closeness and degrees were conducted to provide a greater insight into the dynamics of the formed relationships. These SNA measures are commonly used to determine an individual’s centrality in the network (Otte & Rousseau, 2002). Haythornthwaite (2001) maintains that measures of centrality illustrate “how well positioned an individual is to receive and disseminate information” (p. 216) among the social network. For example betweenness refers to the frequency an individual occurs within the shortest path between other nodes (actors) (Otte & Rousseau, 2002). Participants exhibiting a high betweenness value are often referred to as gatekeepers or brokers as a result of their controlling position in influencing the flow of information and resources in the network (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2001). The measure of an individual actor’s closeness in the network is taken as a more general indicator of centrality. Closeness is defined as the degree of relationship an actor has with the entire network. For example, an actor with a high closeness value will demonstrate
a linkage to other actors through a small number of paths (Otte & Rousseau, 2002). The final SNA calculation conducted was the determination of an individual’s number of ties with other actors. The degree centrality is the number of connections each actor possesses in the network (Otte & Rousseau, 2002). Figure 2 illustrates an example sociogram generated from the present study. I – represents a high closeness centrality, in this instance ‘I’ is the academic staff member associated with the teaching unit. II - indicates a student with a high betweenness centrality within the network. III – illustrates a student with a high degree centrality. IV represents a student disconnected from the main network structure.

Figure 1. Screen capture of the open source software NetVis 2D

Figure 2. Sociogram of discussion forum interactions
Discussion forum analyses

The analysis of discussion transcripts is a common methodological approach adopted by researchers investigating CMC environments. For example, Gunawardena, Lowe, and Anderson (1997) utilised content analysis techniques to investigate the process of co-constructing knowledge in computer conferencing. Similarly, Hara, Bonk and Angeli (2000) conducted content analyses to identify the cognitive skills and meta-cognitive strategies employed in a student discussion group. The present study analysed the forum communication artefacts to identify the quantity of discussion related to the various dimensions comprising sense of community. As Rovai’s (2002b) CCS is comprised of two constructs; social and learning, the forum contributions (n = 899) were categorised according to the dimensions underpinning this scale.

Discussion forum transcripts from eight teaching units were analysed according to Rovai’s (2002a) definition of classroom community and the time of the posting. Selection of the units for analysis was based on the quantity of discussion forum contributions. As each post made to a discussion forum is time stamped, the analysis of discussion content can be conducted along a time scale. Codification in this time-dependent manner serves to illustrate how the discussion forum content transitions through social and learning orientations during the course of the semester. The unit of analysis adopted in this study comprised a sentence in a discussion forum contribution (Hara et al., 2000). Consequently, the posted messages (n = 899) could potentially demonstrate elements of the various community dimensions. In such cases the message is recorded for each specific dimension exhibited.

The codification of the forum transcripts consisted of reading each posting and classifying the message according to a revised version of Rovai’s (2002a) four dimensions of classroom community – spirit, trust, interaction and learning. Postings made by the teaching team (Facilitator) were listed as an additional dimension. The monitoring and reporting of facilitator posts serves to provide an indication of the learning tasks associated with the discussion forum activities and the types of intervention methods that are incorporated to shift the student body from a social to a more learning oriented discourse. In this context the role of interactions relate more to a linkage between the two constructs of sense of community (social to learning). As such, all posted messages (interactions) can be seen to link the two constructs of classroom community. Hence, the dimensions examined for this study included: spirit, trust, learning and facilitator with the amalgamation of spirit and trust forming the social construct. The codification of the posted messages in this format also corresponds to the factors emerging from the Classroom Community Scale (Rovai, 2002b). Table 1 illustrates the classification schema and associated examples.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Definition</th>
<th>Themes</th>
<th>Examples from content analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spirit</td>
<td>The feeling of belonging and connectedness with other class members</td>
<td>Belonging, support, and reference to other members</td>
<td>“My heart goes out to the …. student you are faced with a larger struggle than most of us. If I can assist you in any way – please let me know”</td>
</tr>
<tr>
<td>Trust</td>
<td>The feeling that class members can be trusted regarding possible feedback and support</td>
<td>Self disclosure, humour, identity building and personal reflection</td>
<td>“Hi all, my name is …. I live out in the middle of nowhere (closest habitation is …)”</td>
</tr>
<tr>
<td>Learning</td>
<td>The notion that knowledge is co-constructed with fellow class members</td>
<td>Critique, opinion, debate, discussion of content</td>
<td>“I am now convinced that everything ‘smacks’ of creativity and that the convergence of artistic and scientific processes to produce ‘anything’ deserves much closer scrutiny now than I was prepared to give in the past.”</td>
</tr>
<tr>
<td>Administrator</td>
<td>Posts by the teaching team associated with the delivery of the unit content and learning activities</td>
<td>Facilitator posts are identified and coded based on the name associated with the contribution</td>
<td>“Wright (2003) says that novelty alone does not mean creativity. When a two year old does a finger painting, this is new to them, but is it creative? Do you agree with Wright? What other writers are you reading on creativity?”</td>
</tr>
</tbody>
</table>

* Social community is comprised of both spirit and trust.
Case study

A case study was undertaken to provide additional clarification of the relationship between a student’s position in the social network and their perceived sense of community. Data derived from the case study included student interviews, CCS data and SNA.

For the case study unit, student responses to the online survey were quantified and sequentially ranked according to the individual’s reported sense of community. Based on this ranking, students representing the highest and lowest CCS scores were selected for individual informal, semi-structured interviews (N = 4). Thus, the interviews were designed to further clarify and inform the quantitative data by investigating the deviant cases (high and low community scores) (Patton, 2002). As such, the interview data are seen as supplementary and complementary to the previous data sets. As the teaching unit was offered via an external modality, access to individual students was limited to telephone interviews.

All student interviews were conducted at the end of the teaching semester with each interview lasting approximately one hour. The interview questions focused on clarifying the social, educational and environmental reasons for each student’s individual community score. Students were asked to comment on certain aspects of the unit relating to their perceived level of community and the degree of additional socialisation and contact with class members and other peers associated with their individual networks. For example: How many different students did you interact with in this unit? What did you mostly discuss? Were there other individuals or groups outside of the teaching unit that you had contact with for support? Students were requested to elaborate on any additional themes that were of interest that may not have arisen through the general discussion such as professional networks, competition for assessment or resources, and the perceived value of the implemented online resources.

Statistical analyses

Data derived from the study were analysed using the software package SPSS for Windows © (Vers 12.0.1). Statistical analyses incorporated ordinary least squares regression analyses and basic descriptives.

Limitations of the study

The investigation of community dynamics is a complex study influenced by numerous external factors. Consequently, the research design and subsequent analyses have a few potential limitations that may impact on the broader derived conclusions. The generalisability of the conclusions is limited as a result of the study being confined to a single institution. However, the research methodology adopted confirms the CCS scale and the applicability for SNA to inform educators of the design of implemented learning activities. It is envisaged that the results emerging from this study will provide a benchmark for future comparative analyses.

An additional limitation of the study resides in the unknown number of communication exchanges undertaken by the sampled population external to the monitored online environment. This study investigated discussion forum artifacts as a measure of the social network formed among the student cohort. While the discussion forums are the primary mode of communication among students and educators in the institution, additional communication exchanges are likely to have occurred.

Results

Participants

The overall response rate for the CCS survey was 23% (N = 464) of all students enrolled in the sampled units. Although the response rate appears to be relatively low, the survey participants demonstrated a comparable demographic profile with the faculty at large (Table 2). Males represented 16.16% and females 83.84% of the study participants and the majority of the study participants were enrolled in full time study (87%) via an internal modality (80%). The mean age of the study participants was 26.20 (SD = 8.0). The sampled population also undertook an
average of 16.19 hrs (SD = 11.9) of paid employment per week. Table 3 illustrates the mean CCS score and the constructs comprising the CCS for the participating student cohort. Community scores range from a maximum of 80 to a minimum of 0. The mean CCS score for the internal student cohort was 49.0, while the external demonstrated a mean CCS score of 41.8.

<table>
<thead>
<tr>
<th>Table 2. Demographic profile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Survey response</td>
</tr>
<tr>
<td>Faculty student demographics</td>
</tr>
</tbody>
</table>

*Results derived from the University’s census data for 2005

<table>
<thead>
<tr>
<th>Table 3. Mean sense of community scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean a</td>
</tr>
<tr>
<td>(SD = 11.0)</td>
</tr>
<tr>
<td>Internal Students Mean b</td>
</tr>
<tr>
<td>(SD = 10.2)</td>
</tr>
<tr>
<td>External Students Mean c</td>
</tr>
<tr>
<td>(SD = 12.5)</td>
</tr>
</tbody>
</table>

*Community is equal to the sum of the two constructs: social community and learning. Community scores range from a maximum of 80 to a minimum of 0.

a N = 464, b n = 372, c n = 92

Predictors of student sense of community

An ordinary least squares (OLS) regression procedure was conducted with measures of social network analysis as the predictor and CCS as the criterion variables. The adjusted R² value with all measures of SNA was 0.253 with an associated F of 13.21, indicating that a significant yet moderate proportion of the variance in community was accounted for by the measured variables (Table 4). The adjusted R² value for the social community and learning community sub-scales was .23 (F = 11.51) and .21 (F = 10.26) respectively (Table 4). While closeness and degrees were positive predictors of community and its associated sub-scales, betweenness indicated a negative correlation.

<table>
<thead>
<tr>
<th>Table 4. OLS regression of student sense of community on all SNA variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Betweenness</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Closeness</td>
</tr>
<tr>
<td>Degrees</td>
</tr>
</tbody>
</table>

***Correlation is significant at the .001 level
**Correlation is significant at the .01 level

Discussion forum analysis

Content analysis of the forum postings (n = 899) revealed that during the first 6 weeks of the teaching period social postings are the dominant interactions (Figure 3). Learning interactions demonstrate an increase from 28% to 45% of the communication interactions for the entire teaching period (Figure 3). Facilitator posts appear to influence the level of learning oriented discussions undertaken. For example in week 5 there is a substantial increase in the percentage of facilitator posts – a subsequent increase in learning interactions follows for the week 6 teaching period.
A similar trend is observed in week 7 with an increase in facilitator posts and a subsequent and more sustained rise in learning interactions.

![Figure 3. Percentages of CCS constructs of the total interactions](image)

**Student interviews**

Comments and data deriving from the student interviews (n = 4) is located within the context of the discussion to illustrate examples and augment the interpretation of the quantitative and qualitative findings. Interview comments are presented as Students A and B reporting high CCS scores and Students C and D reporting low CCs scores (Table 5).

<table>
<thead>
<tr>
<th></th>
<th>Community</th>
<th>Social community</th>
<th>Learning community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A</td>
<td>64</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>Student B</td>
<td>64</td>
<td>27</td>
<td>37</td>
</tr>
<tr>
<td>Student C</td>
<td>39</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Student D</td>
<td>33</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Class mean&lt;sup&gt;a&lt;/sup&gt;</td>
<td>48.4 (SD = 8.9)</td>
<td>21.7 (SD = 4.5)</td>
<td>26.7 (SD = 6.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup> n = 13

**Discussion**

This study investigated the relationship between an individual’s position within the social network and their overall perceived sense of community. The findings demonstrate that a relationship exists between student sense of community and the position within the formed social network. The results indicate that the SNA centrality measures of closeness and degrees are positive predictors of an individual’s perceived sense of community. In contrast the SNA measure betweenness, demonstrated a negative correlation with sense of community (Table 4). Furthermore, qualitative analyses indicate that an individual’s pre-existing social network influences the type of support and information exchanges an individual requires and therefore, the degree of sense of community ultimately experienced.
A common theme emerging from the literature relating to community studies is the requirement for social interactions to occur in a regular and timely fashion in order to foster a sense of community among members (Palloff & Pratt, 1999; Wood & Smith, 2005). Although this may appear obvious there has been minimal large scale empirical research undertaken to substantiate this claim. Dawson (2006a) empirically demonstrates that not only is the quantity of communication episodes an important driver for community development but the types of interactions that occur within the online learning environment also influence a student’s sense of community. The author maintained that discussion forums exhibiting a high level of learner-to-learner interactions demonstrate a greater student perceived sense of community.

The closeness and degree centrality measures conducted in the present study also illustrate that students engaged with a greater number of learners report a higher level of sense of community than their less socially active peers (Table 4). The degree and form of reciprocation an individual receives among a group of learners may be an influencing factor in the degree of community an individual experiences (Dawson, 2006a, 2006b). For example Vonderwell (2003), suggests that students contributing to a class discussion forum experience a degree of frustration when their messages are not reciprocated. The author also noted that an individual’s specific learning requirements impact on their perception of the overall group cohesion and the effectiveness of the CMC environment to facilitate learning. Thus, the monitoring of an individual’s evolving social network may act as a lead indicator of sense of community by informing practitioners of an individual’s learning support requirements. For example, students demonstrating a high closeness or degree centrality may be co-located in tightly formed cliques. Hence, in this instance, students may already be receiving the forms of support and information exchanges necessary for their academic progression through the course. In contrast, students unconnected to the network may be experiencing a degree of frustration as a lack of response to their initial inquiry or statement. In this case, the isolated student requires additional bridging support to assist them in linking back to the main internal social network (See figure 2 – IV). Through active monitoring of a student’s level of interaction and position in the social network practitioners are able to offer a more personalised education experience.

Cho, Stefanone and Gay (2002) have noted that the types of information and how information is shared within the CMC setting is influenced by the network structure and the interactions in which a student participates. Thus, the quantity and quality of the communication exchanges conducted can be seen to be largely influenced by an individual’s personal learning and social requirements. Students exhibiting a strong content knowledge may be less inclined to engage in early elementary disciplinary exchanges. Conversely, students with limited external networks may seek a greater social attachment with the internal network and therefore invest a greater proportion of their available time in discussing a broad range of unit related and social content. Social capital is a theory that can be used to inform why students choose to invest, or not invest, in particular discussion forum activities and therefore shape the form of content and topics the cohort discusses and the manner in which the information flows through the network.

**Social capital**

Research on social capital has broadly focussed on the notion that individuals with diverse social networks have the potential to draw upon and generate a greater level of social capital (Cho et al., 2005). In the education context, students have the opportunity to access the new developing learning network in addition to any established pre-existing networks outside of the immediate environment. As such the education setting forms only a subset of the student’s broader social system.

The betweenness centrality data observed in this study suggests that students occupying these positions act as bridges between previously separate cliques. Students exhibiting high betweenness scores also demonstrate high levels of interaction. The students appear to be investing substantial resources into the internal social network structure. However, the CCS scores observed for students with a high betweenness value, suggests that the expected returns did not match the overall investment. This might be a result of students exhibiting a high betweenness value accessing an effective pre-existing external social network in comparison to other class members. As such, students occupying these ‘brokerage’ positions may be seeking very specific supplements to their existing social capital. Consequently, these individuals might require a more specialised level of support that is not realised through the general communicative exchanges with other learners.
The quantity and quality of resources and information exchanged among the student cohort has the potential to influence their experience and hence willingness to further invest in this social organisation. Analysis of the data derived from the student interviews revealed that individuals reporting a high degree of sense of community had limited external (to the institution) contacts that could assist and discuss the unit content and review forthcoming assessment items. For example, interviewed students were asked if time spent reading and participating in the unit discussion forum was beneficial to their learning.

Student A (high CCS score): Yes very. I live in an isolated country town and sometimes there is no adult interaction, the discussion forum gave me some sanity and connection with other adults.
Student B (high CCS score): Yeah.. I could make sure I was on-track and my thinking about the assignments was on the same lines as the others.
Student C (low CCS score): Yeah – it was ok if I needed more I would have done more.
Student D (low CCS score): Not really, the comments were a bit mushy.

The comments noted above illustrate that the level of commitment and satisfaction derived from the participation in the online learning environment is influenced by the perceived benefit of student discussions. Comments made by Students C and D indicate a low commitment and level of satisfaction with the learning community.

When students were questioned about the level of additional external networks for academic and personal support utilised in completing the unit, Students A and B reported having limited external contact in contrast to Students C and D. Student A noted that additional external contacts were limited to a single family member, currently residing inter-state, and an international acquaintance. Consequently, Student A was extremely reliant on the relationships formed through the unit discussion forum in order to access additional information and materials that may assist in their academic completion. Student B, similarly indicated that external contacts were limited. This student referred to a “few work colleagues” as external contacts providing additional support, however, conversations relating to study in these circumstances were noted to be infrequent. In these instances, Students A and B had limited additional social capital to draw upon to aid in the completion of the unit tasks and as such invested considerable effort in developing and maintaining social relations with fellow class members. Conversely Students C and D both noted they had access to a high number of additional external relationships in order to discuss unit content and forth coming assessment tasks. For example, Student D stated that additional external support was frequent (fortnightly) and specific to their current area of study (work colleagues in early childhood professions). In this context, students C and D can be seen to be attempting to supplement their potential external resources via the unit discussion forum. Thus, the level and type of support a student requires becomes increasingly specialised and the ongoing lack of fulfilment results in the individual potentially considering two options. Firstly, students may opt to re-invest further in the relationship in an attempt to satisfy their expectations and requirements or alternatively, the individual abandons the relationship and re-establishes and invests in existing external social structures.

The high degree of betweenness observed in Student D is reflective of this individual re-investing in establishing social connections through the unit discussion forum. Examination of the forum postings related to this individual and the subsequent replies illustrates a mismatch between the social and learning community. Comments posted to the forum by Student D reflect a learning discourse however, subsequent replies reflect a more social and supportive discourse. Reflecting on this in the interview Student D commented:

*The discussion weren’t very (pause) the questions posted were too obscure to answer practically. The comments were all very personal and soul searching. There wasn’t much education value or theoretical veracity. Too many students were talking about their children being creative - it was all very personal.*

This mismatch between investment and return impacted upon the student’s overall level of community experienced.

### Social ties

Numerous authors have commented on the social dynamics between weak and strong ties and the degree of resource exchange occurring in a network (Granovetter, 1982; Haythornthwaite & Wellman, 1998; Marsden & Campbell, 1984). Essentially actors connected via weak social ties are less likely to share resources in contrast to actors connected via strong ties. Within a learning domain the development of strong ties can therefore be seen to be a goal in order to facilitate sharing of resources and information and extending this further, the co-construction of
knowledge. Analysis of the data emerging from this study indicates that the social connections formed among the student members were predominantly weak ties. For instance the overall level of sense of community reported among the student cohort appear relatively low (mean = 47.5, see table 3). Distilling the cohort into both internal and external students revealed that students undertaking the distance education program reported even lower scores (mean = 41.8, see table 3).

The codification of discussion forum interactions along a time-line (Figure 2) illustrates the consistency of the social community for the duration of the teaching period. The observed reliance on the social discussion (e.g. messages re-confirming a sense of belonging and instances of social support) could be interpreted as indicative of a community that is composed of weak social ties. As the community progresses towards a more learning oriented discourse the occurrence of students disagreeing, debating and offering contrary opinions increases. Early literature investigating asynchronous computer-mediated communication often described the medium as devoid of social cues (Sproull & Keisler, 1986). Although authors such as Rheingold (1994) highlight the benefits of this anonymity to lower social inhibitions and generate more candid conversation there is an associated increase in the risk of flaming or misinterpretation which may lead to a diminishing of trust and sense of community (Daft & Lengel, 1986).

Walther (1992) argues that the CMC environment can be as rich in social cues as offline settings given sufficient time and interaction among members. However, within the education context there is a restricted time frame to form strong social bonds among members. Figure 4 illustrates a conceptual interaction pattern occurring within an educational discussion forum. In this instance, the limited degree of social interactions is used merely to supplement the learning interactions which can be seen to be the dominant discourse. The formation of this trend is more aligned with a community formed on strong social ties whereby debate and opinion are presented without affecting the trust and sense of belonging previously established through the socialisation phase. This conceptual framework is akin to Salmon’s (2000) 5 step model of e-moderating. Salmon (2000) poses a methodology for online facilitators to progress a cohort through a socialisation phase (steps 1 and 2) towards discussions centred on the co-construction of knowledge (steps 3-5). The difficulty arising in the adoption of this model is the rapidity at which a cohort needs to progress through the socialisation phase in order to establish strong ties that scaffold future discussions of unit content.

The development of the specific learning activities that promote the conceptual trend are influenced by the size of the cohort and the time restrictions associated with the administration of higher education courses. Schellens and Valcke (2006) codified 38 discussion groups to identify instances of knowledge construction. The authors note that in smaller groups there is a greater proportion of discussion devoted to knowledge construction. The findings suggest that the time required to form a strong social tie among community members is reduced as result of the small group size.
Practical implication: Role of the facilitator

A recommendation arising from the present study is to open the CMC environment to the student cohort prior to the official course commencement date in order to promote early socialisation. It is theorised that the shift to a learning focussed discourse will therefore, occur earlier in the course allowing for an increased opportunity for students to co-construct knowledge within a learning community.

A further recommendation is to promote and monitor facilitator input to assist in the transition of community members from social to the learning focussed discourse. Northedge (2003) has argued that the adoption of constructivist principles in higher education has resulted in a pedagogical shift from didactic teacher to facilitator (sage on the stage to guide on the side). The author further noted that this adoption is not necessarily indicative of sound learning practices. Northedge (2003) stated that a balance between “delivering knowledge” and more student-centred practices is necessary. Hence, a conscious and purposeful learning intervention is often required to shift student discourse from the social to a learning orientation. McWilliam (2005 p. 5) describes this teaching role as the “meddler in the middle”. Considering the paucity of social cues and the necessity for developing a social presence in an online education environment active and substantial early involvement by academic staff can be seen to be critical for assisting students in developing a strong and effective learning network.

Conclusion

This paper presents the findings from a study investigating the relationship between a student’s position in the social network and their perceived sense of community. The findings suggest that the position an individual occupies in the social network is indicative of both their degree of perceived sense community and the nature of the academic and social support the individual requires for future progression through the course. Furthermore, the qualitative analyses illustrate that the time required to transition student discourse from a social nature towards a learning orientation forms a large proportion of the available teaching period. Therefore, additional socialisation activities are required to both facilitate the rapid formation of social relationships and to enhance the overall strength of the social ties formed.

Contemporary government and education policy continue to emphasis the requirement for institutions to demonstrate the provision of quality outcomes and practices. The integration of SNA measures with the institutionally adopted LMS can be utilised to inform practitioners of the degree of personalised support an individual student requires. Further research is required to investigate the relationship between student social networks and other dimensions influencing the student learning environment such as the specific pedagogy employed, practitioner personality and the cohort demographic profiles.

References


Modeling and Intervening Across Time in Scientific Inquiry Exploratory Learning Environment

Choo-Yee Ting\(^a\), Somnuk Phon-Amnuaisuk\(^b\)

Faculty of Information Technology, Multimedia University, Cyberjaya, Malaysia // \(^a\)cyting@mmu.edu.my // \(^b\)somnuk@mmu.edu.my

Yen-Kuan Chong

Center for Multimedia Education Development & Application, Multimedia University, Cyberjaya, Malaysia // ykchong@mmu.edu.my

ABSTRACT

This article aims at discussing how Dynamic Decision Network (DDN) can be employed to tackle the challenges in modeling temporally variable scientific inquiry skills and provision of adaptive pedagogical interventions in INQPRO, a scientific inquiry exploratory learning environment for learning O’level Physics. We begin with an overview of INQPRO and a highlight of the computer algorithm as well as the design of our proposed DDN model. We then present an instance of interactions with INQPRO to describe how the proposed model can be generated dynamically by aggregating different INQPRO Graphical User Interfaces (GUIs) in real-time basis to perform probabilistic assessments of the two scientific inquiry skills (Hypothesis Formulation Η and Variable Identification ζ). In this study, we carried out a two-phase empirical evaluation to investigate the performance of the proposed DDN model in categorizing different groups of learners. The performance of the proposed DDN model is identified by its matching accuracies elicited from a total of 6 domain experts and 77 learners who participated in both evaluation phases. Based on the empirical results, we summarized that the proposed DDN model is practically sound as it has demonstrated acceptable estimation accuracies with reference to the classification results obtained from the pretest, posttest, and from domain experts.

Keywords

Dynamic Decision Networks, Scientific inquiry learning environment, Learner modeling.

Introduction

Researchers in the field of science education have been employing scientific inquiry as an instructional strategy to maximize learner's engagement and experiences during the learning processes (Frederiksen & White, 1998; Hulshof & de Jong, 2006; Linn, 2000; Pryor & Soloway, 1997; Reiser et al., 2001; Shimoda, While & Frederiksen, 2002; Veermans & van Joolingen, 2004). The importance of this instructional strategy is proven by the increasing number of computer-assisted learning environments developed recently such as the Belvedere (Frederiksen & White, 1998), BGuilLE (Pryor & Soloway, 1997), KIE (Linn, 2000), SCI-WISE (Reiser et al., 2001), SimQuest (Veermans & van Joolingen, 2004), Rashi (Dragon et al., 2006), and SmithTown (Shute & Glaser, 1990). These learning environments do not directly deliver scientific facts to learners. Instead, learners are required to actively involve in scientific inquiry processes such as evidence gathering, constructing and testing hypotheses, manipulating variables, and the like. As an attempt to maximize scientific inquiry learning experience, exploratory learning approach is often the preference when it comes to development of computer-based learning environments (Chang et al., 2003; de Jong & Van Joolingen, 1998; de Jong, 2006). The exploratory learning approach provides learners with freedom to interact with the learning environment by means of test-and-retest their idea.

To date, although there has been an attempt to integrate intelligence into scientific inquiry learning environment by employing a learner model (Meyer et al., 1999), solutions to the following challenges remain unclear. Firstly, how should a learner model be integrated into a scientific inquiry learning environment that consists of more than a single GUI and rooted in a particular instructional model? Secondly, having gathered a learner’s exploratory behaviours, how should a system effectively assess the mastery levels of scientific inquiry skills which evolve across time? Thirdly, how should a system generate tailored pedagogical interventions in a timely manner to cater for temporally variable scientific inquiry skills? These challenges are not trivial as the system has to deal with a high level of...
uncertainty inherent in inferring a learner’s mastery level of scientific inquiry skills from exploratory behaviours (Schum, 1994; de Jong, 2006).

To handle the uncertainty in an efficient modeling of interaction behaviours, Bayesian Network (BN) (Pearl, 1998) has been employed as one of the solutions (Jameson, 1995). However, a BN does not provide decision making under uncertainty. As a complementary solution, a Decision Network (DN) (Howard & Matheson, 1981; Russell & Norvig, 1995), which is an extension of BN with decision and utility nodes, is proposed by researchers in the field of Artificial Intelligence in Education (e.g., Murray et al., 2004; Conati, 2002, Pek & Poh, 2005). When time is a crucial factor in learner modeling, a static DN is then extended to a Dynamic Decision Network (DDN). By employing a DDN, not only the system is able to model the variables that evolve across time, but at the same time provides tailored feedback in a timely manner. A study by Murray et al. (2004) has shown that the decision-theoretic approach outperforms Fixed-Policy Tutor in selecting the optimal tutorial action. Here, we review instances of decision-theoretic Intelligent Tutoring Systems (ITSs) such as DT Tutor (Murray et al., 2004), CAPIT (Mayo & Mitrovic, 2001), Prime Climb (Conati, 2002), and iTutor (Pek & Poh, 2005), and subsequently highlight the major differences between their work and ours.

DT Tutor is an ITS designed for two domains, namely, the Calculus-related rate problems and Elementary reading. DT Tutor takes into consideration of not only the learner’s goals, focus of attention, and affective state, but also its objectives to provide optimal pedagogical decision. DT Tutor employs a DDN for selecting tutorial actions while assisting a learner with a task. DT Tutor was evaluated to verify if the selected tutorial action is rational and fast enough under a variety of tutorial conditions through an extensive simulation. Both DT Tutor and our work decide optimal tutorial action with maximum expected utility and exploit temporal learner properties. Although both DT Tutor and our work rely on DDN to model and intervene under uncertainty, our work differ from DT Tutor in several ways. First, a DDN is leveraged by DT Tutor to predict multiple steps ahead, while ours do not. The difference is largely because of the nature of INQPRO (this research work). INQPRO is based on exploratory learning approach (de Jong, 2006) which consists of six different interfaces. Each interface in INQPRO has a DN associated with it. Considering the number of nodes in each DN and all the possible navigation paths (state space = 5^n with n ∈ {Integer > 0}), predicting multiple steps ahead can easily be computationally intractable. Second, the emphasis of DT Tutor is studying how to recognize the evolving learner’s affective states, and focus of attention; while our work focuses on both how to recognize learner’s temporally variable scientific inquiry skills and on how learner assistance can be timely and appropriately applied within the decision-theoretic framework. Instead of merely developing a component for optimal pedagogical intervention selection mechanism, our work involves developing the interface component and the inference engine as well as their systematic integration into INQPRO. Third, the effectiveness of the DT Tutor’s tutorial action section engine is evaluated via an extensive simulation. INQPRO, however, is evaluated and validated via simulation, human learners, and domain experts.

Prime Climb is an ITS that enhances a learner’s acquisition on Prime numbers. By interacting with the learner, the Intelligent Pedagogical Agent monitor the learner’s emotions and generate appropriate interventions aiming at achieving the best tradeoff between the learner’s learning and engagement during their interactions with Prime Climb. To achieve this objective, the intervention is decided by maximizing the expected utility of all available decision alternatives. The DDN employed in Prime Climb is closely related to that implemented by DT Tutor, where the ‘intervening’ type (Jensen, 2002) of DDN is employed. The difference between the DDN employed by DT Tutor and that of Prime Climb is the number of look-ahead time-slices generated. Different from Prime Climb, however, there is no look-ahead time-slice due to the computational complexity in evaluating the DNs.

CAPIT, a Constraint-Based Tutor for teaching capitalization and punctuation, is an ITS that leverages decision-theoretic approach to decision making. The major difference between CAPIT, DT Tutor, Prime Climb, and our work is that CAPIT separates the learner modeling portion from the tutorial action selection engine. It has a BN that performs assessment on a learner’s knowledge states. Based on the outcome probabilities computed by the network, they are then multiplied with their associated utilities to determine which decision alternative has the maximum expected utility. Such calculation, however, is performed outside the network itself. Performing calculation of expected utilities outside a BN is deemed to potentially faster inference as decision and utility nodes are removed. According to Murray et al. (2004), such potential speedup of inference might resort to two drawbacks. First, there is a possibility that less obvious decision alternatives are not taken into consideration. Second, computation of expected utility values for each decision alternative might not be done accurately.
iTutor employs a DDN to pre-compute a tutoring policy such that the tutoring action that maximizes the expected utility. The optimal decision may belong to one of the three main categories: (i) assessing the learner’s level of knowledge mastery, (ii) presenting a lesson, and (iii) determining the learner’s readiness to leave the tutoring session. If the decision is to assess the student’s mastery, a challenging item will be selected. If the decision is to present a lesson, iTutor will select a suitable instruction from its domain knowledge database. While iTutor focuses more towards the physics concepts acquisition, our work focuses on mastering of scientific inquiry skills. Besides, rather than pre-computing a tutoring policy, our work focuses on localized and fading support as the learner interact with INQPRO. We shall further discuss the localized and fading support in a later part of this article.

The ITSs mentioned above share a common feature: the DDNs employed are having a set of nodes that are identical for all the \( n \) time-slices. Different from our work, the DDNs employed by these ITSs do not contain static node. Such approach, however, is not applicable generally. For instance in study, the INQPRO consists of six GUIs and there is a DN associated to each GUI. Thus, the main challenge in this study is to identify a sound DDN model that is able to handle the time factor gracefully without affecting its accuracy in assessing learner’s scientific inquiry skills.

In the rest of this article, we shall firstly provide an overview of DDN and the INQPRO learning environment. We then proceed with the discussion on the detailed design of our proposed DDN model. To provide a better understanding of how the proposed DDN model functions, an illustration on the probabilistic assessment of scientific inquiry skills and generating tailored interventions is given. Lastly, we shall discuss the empirical results from the evaluation of the proposed DDN through field tests administrated on human learners and classifications elicited by human experts.

**Dynamic Decision Network**

Figure 1(a) depicts a BN with three chance nodes while 1(b) depicts the extension of BN into a DN by adding utility (\( Y \)) and decision (\( A \)) nodes. A DN, which is also known as Influence Diagram (Howard and Matheson, 1981; Jensen, 2002), is extended from a BN to allow decision making under uncertainty. Similar to a BN, the formalism of DN allows encoding for probability distribution over a set of random variables. On top of that, it provides the capability to present decision alternatives that an agent can take together with the utilities of the uncertain outcomes. The qualitative perspective of a DN is a graphical representation of the variables that takes the form of a Directed Acyclic Graph (DAG).

![Figure 1. (a) a BN; (b) a DN](image)

The three types of node found in a DN are the chance node, decision node, and utility node. The chance node in DN is similar to that in BN. It represents a random variable and usually represented using a circle (nodes \( X_1, X_2, X_3 \) in Figure 1). It has an associated Conditional Probability Table (CPT), giving the probability of the variable given its parents. The parent nodes of a chance node can be both chance nodes and decision nodes. The arcs between the chance nodes encode the interdependences among the related random variables. The second type of node is the decision node (node A in Figure 1(b)). It represents all the available decision alternatives being taken at a particular time point and usually depicted as a rectangular shape. The values of a decision node represent the decision alternatives to be chosen between. To decide among the decision alternatives, the expected utility of each alternative is calculated by summing the utilities of all possible outcomes weighted by the probabilities of those outcomes:

\[
EV(A|\pi) = \sum_{\text{Outcome}_i} P(\text{Outcome}_i|\text{A}) \cdot U(\text{Outcome}_i|\text{A})
\]

(Eq. 1)
From Equation (1), action $A$ has possible outcome states $\text{Outcome}(A)$, which is the current knowledge (evidence) of agent. The agent assigns probability $\mathbb{P}(\text{Outcome}(A) | \text{evidence}(A))$ prior to execution. $\text{Action}(A)$ is the proposition that action $A$ is executed in the current state. Following the principle of Maximum Expected Utility (MEU), a rational agent should then select an action that maximizes the agent’s expected utility. That is, to maximize Equation 1.

A decision node can have both chance nodes and other decision nodes to be the parent nodes. Having another decision node as parent node, the network is actually modeling a sequence of decisions to be taken. The third type of node is the utility node $\Upsilon$. It highlights the desirability of the consequences that may arise from the various decisions. Utility nodes are depicted as a diamond. It can have both chance nodes and decision nodes as parent nodes. The incoming arc of a utility node represents an influence on preference. However, there is no outgoing arc from utility nodes. The arc directed from a decision node to a chance node represent an influence on random variable exerted through an intervention from the decision alternatives at hand.

Figure 2. A general three time-slices DDN. In this study, this type of DDN has been integrated and evaluated in the INQPRO system. There is no temporal dependency between the decision nodes

A DN in Figure 1(b) can be extended with a temporal dimension resulting in a DDN (Figure 2). A DDN is a graphical data structure that models the state of the world over time and typically represents a particular number of connected time-slices. The relationships between variables at successive time steps are represented by temporal arcs. The variables that evolve across time are considered as dynamic nodes while variables that are instantiated in a particular time-slice only are known as the temporal nodes (Schafer & Weyrath, 1997). In some instances of DDN, there are variables that do not evolve across time. Such variables are commonly expressed as the Static nodes. The DDN structure employed in this study does not have temporal arcs that span more than a single time step. This is an instance of the Markov Assumption, which the state of the world at a particular time depends only on the previous state and any decision taken in it. The relationships between the variables are quantified by the conditional probability distribution associated with each node. Once the DDN is defined, the agent’s current beliefs about the world are set as evidence in the time-slice $t_n$. By updating the DDN, the expected utility of performing each possible decision can be calculated and the agent can select the decision alternative that has the highest expected utility.

In this study, the DDN created is following the Markov assumption that the state of the world at a particular time depends solely on the immediate previous state and any action taken in it. Thus, there are no arcs that span more than a single time step. In the subsequent sections, we describe in detail the sample assessment of learner’s hypothesis formulation and variable identification skills using a scenario.

An Overview of INQPRO Learning Environment

INQPRO is a computer-based scientific inquiry learning environment for enhancing scientific inquiry skills acquisition. The development of INQPRO is rooted in Scientific Inquiry Exploratory Learning Model (Ting & M.Y. Arshad, 2000) (Figure 3). This model was employed mainly because it emphasizes on the explicit instructional steps that form the basis of the INQPRO development. The GUIs in INQPRO consist of the Scenario (Sc), Hypothesis Visualization (Vz), Verification (Vf), Formula (Fe), Simulation Experiment (Ex), Data Comparison (Dc), and
Feedback (Fz). The GUIs allow the scientific inquiry skills (Hypothesis Formulation and Testing H and Variables Identification ζ) to be assessed separately (within a particular GUI) or as a whole (overall acquisition level after n times of interactions). By actively interacting with the GUIs and Animated Pedagogical Agent (hereafter “Agent”), learners are ultimately expected to command two scientific inquiry skills: H (node H in Figure 4) and ζ (node ζ in Figure 4). Since H and ζ influenced each other, we introduced K (node K in Figure 4) during the modeling phase to represent the overall acquisition of both H and ζ. This would allow us to map K with the results of pretest and posttest, which formed the basis for computing the matching accuracies. To maximize learning experiences throughout the learning process, we adopted exploratory learning approach to furnish learners with the freedom to explore the domain concepts through manipulation of the scientific inquiry skills. This can be achieved by integrating rich interactive learner control components such as the drop down box, list box, track bar, and drag-drop objects in the GUIs.

Figure 3. The Scientific Inquiry Learning Model. The names of INQPRO’s GUIs are given on the left pane. Each GUI has its own DN.

Figure 4 depicts the high-level presentation of INQPRO’s DNs. The network is categorized into four sub-networks with each sub-network consisting of either observable or non observable nodes. The observable nodes (nodes with prefix of SA_, SQ_, AQ_, t ) are nodes to be instantiated in light of evidence. In this context, learner’s interactions such as the keystrokes, mouse-click, drag-and-drop, and the typing events are instances of evidence. This kind of activities provides an indicator to how a learner reasons and making sense of H and ζ. Examples of observable nodes in Figure 5(b) are nodes AQ_Concept, SQ_Definition, and AQ_Scenario. The SQ_Definition for instance, aims to capture the frequency of questions asked by a learner. Once evidence is obtained and corresponding nodes are instantiated, interpreting this information into a DN allows INQPRO to suggest the mastery levels of K, H and ζ and subsequently trigger the Agent to intervene. Thus, the more learner interactions are captured, the higher the bandwidth (VanLehn & Martin, 1988). This subsequently helps reduce the uncertainty in modeling the mastery levels of K, H and ζ as they interact with INQPRO. Conversely, the non observable nodes are nodes that cannot be instantiated directly; they can only be inferred once observable nodes are instantiated and DN is updated. Examples of non observable nodes in Figure 5(b) are the nodes Hypothesis and Variable. Each DN consists of at least one decision node A (nodes prefix with AA_). The decision node in the Scientific Inquiry subnetwork for instance is meant for the Agent to provide tailored pedagogical intervention to foster the acquisition of H and ζ.
To enable the modeling of a learner’s $H$ and $\varsigma$ at a particular GUI, a DN needs to perform both diagnostic and predictive reasoning. By performing diagnostic reasoning, a learner’s $H$ can be inferred from the following evidences: (i) the correctness of hypothesis statement (node $SA_{\text{HypoStruct}}$), (ii) the correctness of variable relationships statement (node $SA_{\text{HypoRelation}}$), and (iii) whether or not hint is requested from the agent (node $SA_{\text{AskHypo}}$). The predictive reasoning, conversely, offers an indirect assessment through the propagation of probability from nodes $H$ and $V$ to node $K$. Detailed explanations regarding the design and integration of nodes into the GUIs can be found in our previous work (Ting et al., 2006).

In the next section, we shall provide a high-level illustration and discussion of two GUIs from the INQPRO learning environment namely the Scenario and Hypothesis Visualization. We present how the corresponding DNs is integrated into the GUI to allow reasoning and intervening under uncertainty.

The Scenario GUI

Figure 5 depicts the Scenario GUI of INQPRO. It is designed by taking the first two phases (Scenario Presentation and Hypothesis Formation, Variables and Relationship Identification) from the learning model (Figure 3). Upon logging into the Scenario GUI, learners are requested to select and read a scenario (Figure 5-c). Once a scenario is selected, a three dimensional computer animation is presented to the learner. The computer simulation acts as an advance organizer (Ausubel, 1968) aiming at presenting a global presentation of knowledge to be learned and foster the acquisition of scientific inquiry skills. Having read the scenario, learners then proceed to the Hypothesis section (Figure 5-d) to construct a hypothesis statement that well describe the scenario. The learner will then be requested to identify the different types of variables in the Variable section (Figure 5-e). INQPRO does not require learners to draw graph, as drawing of a graph by itself is not a trivial task and has been the object of instruction in itself (Karasavvidis et al., 2003). Therefore, by selecting suitable answers for x-axis, y-axis and typing in data sets, graph will be plotted based on the relationship of variables as reflected through the hypothesis formulated. Thus, different hypotheses might result in different patterns of formulated graphs.
The Hypothesis Visualization GUI

Figure 6(b) depicts the DN for Hypotheses Visualization interface. To infer the degree to which a learner is considered to have mastered the variables (node Variable\_Vz) and hypothesis (node Hypothesis\_Vz) relies on whether or not s/he is able to analyze the graph (node AnalyzeGraph) and has understood the purpose of simulation (node UnderstandAnimation). However, both AnalyzeGraph and UnderstandingAnimation cannot be observed directly from the interface. To obtain the posterior probability values for these two nodes, the network performs diagnostic reasoning given the instantiation of evidential nodes (nodes AQ\_MassAni, SA\_PlayAni, SA\_DragMass, SA\_ViewGraph, SA\_CompareGraph, and AQ\_CompareGraph).

In the next section, we shall present the description of a high-level presentation of three time-slices DDN and how the DDN is generated when learners navigate from one interface to another.
A DDN Approach for INQPRO

In this study, we employed DDN to tackle the challenges in assessing temporally variable learner’s scientific inquiry skills for three main reasons. Firstly, a learner’s scientific inquiry skills evolve across time, thus capturing the dependencies between the temporally variable skills is crucial. Secondly, freedom to navigate from one GUI to another introduces a complexity in predetermining a DDN. A predetermined DDN can easily become computationally intractable as it exhibits $5^n$ state spaces (combination of different navigation paths) with $n \in \{\text{positive integer}\}$. Thirdly, if a static DN is employed, the interpretation of new evidences will lead to the reinterpretation of previous evidences. If past learning experiences are not accounted, there is a tendency for INQPRO to inaccurately model the mastery level of scientific inquiry skills and consequently the interventions generated are not tailored to the learner. In order to overcome this drawback, a DDN must be employed instead of a static DN.

Classification of Nodes in DDN

As depicted in Figure 7, time-slice $t_i$ represents the current INQPRO interface accessed by a learner while time-slice $t_{i-1}$ describes the immediate previous state. To describe the immediate subsequent interface accessed, the time-slice $t_{i+1}$ is used. These time-slices are interconnected by temporal relations, which are illustrated by the arcs joining variables that evolve over time.

Figure 7 depicts the high-level presentation of the proposed DDN structure employed in this study. The nodes in the proposed DDN structure are categorized into dynamic nodes, temporary nodes, static nodes, decision nodes, and utility nodes. These four types of nodes share a common format $\text{Node}_{\text{gui}}^{\text{type}}$ with $\text{gui} \in \{\text{All INQPRO GUIs}\}$, while $n$ denotes the frequency a particular GUI is accessed. As an example, given that a learner who accesses the Scenario GUI for twice, then $\text{Node}_{\text{gui}}^{\text{type}}$ represents the first access of Scenario GUI while $\text{Node}_{\text{gui}}^{\text{type}}$ represents the second time the similar GUI is accessed.

In this study, one of the main challenges in this study is how to model a learner’s $\eta$, $\varsigma$, and $\kappa$ that evolve across time. Thus, classifying the nodes $\text{Node}_{\text{gui}}^{\text{type}}$, $\text{Node}_{\text{gui}}^{\text{type}}$, $\text{Node}_{\text{gui}}^{\text{type}}$ into dynamic nodes is appropriate as these nodes are meant for modeling the way the mastery level of $\eta$, $\varsigma$, and $\kappa$ evolve from one time-slice to another. The temporal dependencies are expressed by the arcs (dotted lines) directed between the time-slices (e.g. $\text{Node}_{\text{gui}}^{\text{type}} \rightarrow \text{Node}_{\text{gui}}^{\text{type}}$) in Figure 7. The temporal arcs allow INQPRO to predict the acquisition levels of $\eta$, $\varsigma$, and $\kappa$ at the current time-slice ($t_i$) based on previous experiences instantiated as evidences in previous time-slices ($t_{i-1}$).
The second type of nodes that exist in a particular time-slice only is categorized as the Temporary nodes. These nodes reside in the Interface Interaction subnetwork (Figure 4). Example of nodes that categorized as temporal nodes are the Agent Intervention Nodes (nodes with prefix AQ_) and Learner Exploration Nodes (nodes with prefix SQ_, SA_) The unique property of temporal nodes is that this type of nodes allow the instantiation of evidences at a particular GUI in time slice \( t_i \), not to have the similar evidences instantiation at the similar GUI in time slice \( t_i \). For instance, presumably in time-slice \( t_i \), a learner has typed in a wrong hypothesis statement resulting in the node SA_HypoRelation and SA_HypoStruc to be instantiated to non-mastery. However, revising and refining the hypothesis statement in time-slice \( t_i \) has subsequently resulting in the nodes instantiated to mastery.

Apart from the dynamic nodes and temporal nodes, the nodes \( S^H, S^E, \) and \( S^V \) in Figure 7 are categorized as static nodes. The probabilities of a static node may change but it is not a variable that evolves significantly with respect to each time-slice. Conversely, due to the conditional effect of dynamic nodes (nodes \( P^H_n, P^E_n, P^H_n \) ) on the Static nodes (nodes \( S^H, S^E, \) and \( S^V \) ), a consistent rather than drastic change of probability propagation is observed. This conditional effect provides an “accumulative” probabilistic assessment of \( H, \xi, \) and \( K \) after a series of interactions with INQPRO during the learning process. Such mechanism has provided the DDN with the functionality to remember and consider the mastery level of the immediate previous time-slice when reasoning the mastery levels of \( H, \xi, \) and \( K \) in current time-slices.

A learner who revisits a particular GUI might not need similar coaching as to what one has received during the first visit. This is particularly true when the learner revisits the GUI just to confirm what one has selected or typed in is correct. To handle this challenge gracefully, the agent’s interventions are “localized” to each GUI aiming at capitalizing learners’ learning experiences. Thus, the DDN depicted in Figure 7 has decision nodes (\( D^H_n \)) and utility nodes (\( U^H_n \)) resides in each time-slice. The arcs directed from \( H_n^E, \xi_n^E, \) and \( K_n^E \) to \( H_n^F, \xi_n^F, \) and \( K_n^F \), capturing the idea that the agent’s satisfaction depends on the acquisition level of \( H \) and \( \xi \) and subsequently tailored interventions can be provided at a particular GUI. The configuration encoded in the utility nodes of the DNs in INQPRO will result in the agent’s preference to be “Doing Nothing” whenever there is no evidence instantiation or during a learner’s \( H \) and \( \xi \) are assessed to be “mastery”. Such encoding methodology follows exploratory learning approach where learners are given the freedom to explore, and subsequently make more meaningful learning while interacting with INQPRO. Conversely, if a learner’s \( H \) and \( \xi \) are assessed to be “non mastery”, indirect hint will be prompted accordingly.

Generating DDN during Runtime

Figure 8 depicts the computer algorithm for generating a DDN during runtime. The \( \text{CreateDDN} \) function is called each time a GUI is accessed, which in this context refers to the event when a learner clicks at the “Next” or “Go To...” button. As depicted in Figure 8, the \( \text{CreateDDN} \) function takes into two parameters: \( P \) and \( t \). \( P \) denotes the next GUI to be displayed while \( t \) denotes the \( n^{th} \) time \( P \) is accessed.

The \( \text{CreateDDN} \) function firstly checks whether or not the runtime DDN (\( P_{DDN} \)) exists (refer to line 8, Figure 8). If it does not, a copy of Scenario DN will be duplicated for \( P_{DDN} \). Having set Scenario DN as the first network in \( P_{DDN} \) indicates that it is the first time the learner interacts with INQPRO.

If \( P_{DDN} \) is found, the decision nodes (\( A_{DDN} \)) and utility nodes (\( Y_{DDN} \)) in \( P_{DDN} \) will firstly be removed. Removing these nodes greatly reduces the computational time in updating the DDN. This is particularly crucial as lagging in updating \( P_{DDN} \) will affect the smooth interaction process between learners and INQPRO. The \( \text{AppendNodes} \) function will then update \( P_{DDN} \) with nodes in DN which corresponds to \( P \). The newly added nodes can be differentiated from similar nodes in previous time-slices by the parameter \( t \) which specifies the \( n^{th} \) \( P \) is accessed. Once the nodes in \( P \) are appended to \( P_{DDN} \), the \( \text{DirectArc} \) function create causal arcs directed from the static nodes (\( S^H, S^E, \) and \( S^V \) ) to the dynamic nodes (\( H^F, \xi^F, \) and \( K^F \) ) in \( P \). The \( \text{LoadCPT} \) function defines the Conditional Probability Tables (CPTs) for the dynamic nodes in the current GUI. At this stage of our research work, the Conditional Probability Tables (CPTs) are defined by domain experts and vary from one GUI to another. This indicates that the value of the dynamic node at one time can only affect the value of the similar dynamic node at the next time step depending on the GUIs accessed.
Algorithm 1 CreateDDN($P$, $t$)

Input: $P \in \{\text{GUIs except the Feedback GUI}\}$

$t = \text{the } n\text{th time } P\text{ is accessed}$

$P_{DDN} \leftarrow \text{DDN generated during runtime}$

$\Gamma_{DDN} \leftarrow \text{array of GUIs } \{\Gamma_1, \ldots, \Gamma_n\} \text{ except } P$

$A_{DDN} \leftarrow \text{array of decision nodes } \{A_1, \ldots, A_n\} \text{ in } \Gamma$

$\Upsilon_{DDN} \leftarrow \text{array of utility nodes } \{\Upsilon_1, \ldots, \Upsilon_n\} \text{ in } \Gamma$

If Not Exist $P_{DDN}$ then

$P_{DDN} \leftarrow \text{AppendNodes}(P_{DDN}, Sce)$

$P_{DDN} \leftarrow \text{AppendNodes}(P_{DDN}, \{S^R_1, S^R_2, \text{ and } S^V\})$

Else

$\delta \leftarrow \text{RemoveDecisionNodes}(A)$

$\delta \leftarrow \text{RemoveUtilityNodes}(Y)$

$P_{DDN} \leftarrow \text{AppendNodes}(P_{DDN}, P, t)$

$\text{DirectArc}(\{S^R_1 \rightarrow H_1\}, \{S^V \rightarrow P_1\}, \{S^R_2 \rightarrow H_2\})$

$\text{LoadCPT}(P, t)$

End if

Output: $P_{DDN}$ with $P$ added.

Figure 8. Algorithm for generating runtime DDN

Although the above mentioned algorithm is tailored for INQPRO, it is applicable and transferable to any learning environment consisting of the following three properties:

(i) There is a set of GUIs $G$, such that each GUI $g \in G$ is distinguishable. That is, the learning environment consists of several GUIs and all the GUIs having different interface design.

(ii) There is a set of DNs $D$, such that each DN $d \in D$ is distinguishable, and $g$ has a corresponding $d$. Each $d$ is unique.

(iii) There is an identical set of random variables that evolve across time $\{H_1, \ldots, H_n\}$ for every $d$.

(iv) A DDN is made up of interconnected time-slices of static DNs, where $\{H_1, \ldots, H_n\}$ in every two consecutive time-slices are linked together through temporal arcs. Introduce a set of static nodes $\{S_1, \ldots, S_n\}$ so that arcs can be directed from the static nodes to the corresponding $\{H_1, \ldots, H_n\}$. However, the dependencies between $\{S_1, \ldots, S_n\}$ are domain dependent.

In the next section, we shall present a walk-through with INQPRO in order to provide reader with a better understanding of how the proposed DDN model behaves, which in this context, predicting the learner’s mastery level of scientific inquiry skills and provision of appropriate support in timely manner.

A Sample Walk-Through with INQPRO

The following scenario $P$ aims at illustrating how the proposed DDN model can be employed to assess temporally variable $H$, $\zeta$, and $K$. To simplify the explanation, we shall scope our illustration to only five GUIs accessed by both learners $\Lambda_A$ and $\Lambda_B$ (Figure 9). Let us presume that $\Lambda_A$ denotes a learner whose mastery levels of $H$, $\zeta$, and $K$ have increased while $\Lambda_B$ represents a learner who has failed to achieve meaningful learning after interacting with INQPRO.

Scenario $P$: Supposing that two learners $\Lambda_A$ and $\Lambda_B$ have interacted with INQPRO and both demonstrate identical learning path (Figure 9). In Figure 9, both $\Lambda_A$ and $\Lambda_B$ have started with the Scenario (Sce) GUI and later proceeded to Hypothesis Visualization (Vz), Formula (Fe), Scenario (Sce), and finally the Data Comparison (Dc) GUI. Presumably in time-slice $t_1$, both $\Lambda_A$ and $\Lambda_B$ would have constructed inappropriate hypothesis statements, and unable
to identify correct variables types. However, having performed a series of meaningful interactions (e.g. revisiting the previously formulated hypothesis statement, selecting appropriate variable types, and interacted to the Agent), ΛA has ultimately reconstructed an appropriate hypothesis statement during her revisit to Scenario GUI in time-slice \( t_4 \). Conversely ΛB who did not demonstrate meaningful interactions and has ignored hints generated by the Agent, has failed to reconstruct an appropriate hypothesis statement in time-slice \( t_4 \).

Figure 9. The INQPRO GUIs accessed by both learners ΛA and ΛB. Assuming that after interacting with INQPRO, ΛA is classified as “mastery” while ΛB is classified as “non-mastery”.

Figure 10. The DDN generated after accessing five GUIs. Both DN structures and temporal nodes are different in all the time-slices but are presented in the same structure for ease of presentation.

Figure 10 depicts the run-time generated DDN model that aims at modeling a learner’s \( H \), \( \zeta \), and \( K \). In this model, there are a total of 15 dynamic nodes (e.g. nodes \( H^{t_1} \), \( H^{t_2} \), \( H^{t_3} \), \( H^{t_4} \), \( H^{t_5} \)) and 3 static nodes (nodes \( S^p \), \( S^q \), and \( S^r \)). The static nodes capture the gradual change in the mastery levels of \( H \), \( \zeta \), and \( K \) for the five time-slices through the “accumulative” modeling of the scientific inquiry skills. In the above scenario, given that in time-slice \( t_1 \) that ΛA and ΛB have performed unsuccessful interactions, which include the inappropriateness of constructing hypothesis statement, selecting variable types, and answering questions prompted by the Agent incorrectly, a non-mastery level is suggested by the DDN for both \( H \) (node \( H^{t_1} \)) and \( \zeta \) (node \( \zeta^{t_1} \)). Once the beliefs of the DDN are updated, knowing the posterior probabilities of \( H^{t_1} \) and \( \zeta^{t_1} \), the posterior probabilities of \( S^p \) and \( S^q \) would allow the posterior probabilities of \( S^r \) and \( S^r' \) to be calculated and subsequently the ultimate mastery levels of \( H \) and \( \zeta \) are reviewed. This is possible due to the conditional effects of the static nodes (nodes \( S^p \) and \( S^q \)) on the dynamic nodes (nodes \( H^{t_1} \) and \( \zeta^{t_1} \) as depicted by the arcs directed from the static nodes to dynamic nodes.

Apart from knowing the mastery levels of \( H \) and \( \zeta \), updating the beliefs in time-slice \( t_1 \) has also provided a unique mechanism for the Agent to select the decision alternative (node \( A^{t_1} \)) that maximizes the expected utility (node \( U^{t_1} \)). In the above scenario, the Agent has suggested both ΛA and ΛB to restudy the hypothesis formulated and variables selected in time-slice \( t_1 \) due to the low mastery levels.

Presumably in the Hypothesis Visualization GUI, both ΛA and ΛB have equally interacted with the computer simulations by performing the drag-and-drop event on the different masses, identified different graph patterns generated as a result of different masses, and correctly answered questions prompted by Agent. Once such information is interpreted into the DDN, the mastery levels of \( H \), \( \zeta \), and \( K \) in time-slice \( t_2 \) for both ΛA and ΛB can be
Due to identical interactions elicited by both learners, the mastery levels and probability values are similar in time-slice $t_2$.

![Figure 11](image)

After studying the computer simulations, the learners proceeded to the Formula Investigation GUI. The interaction logs capture the information acquired when $\Lambda_A$ explores the relationships between $m$, $k$, and $T$, and react positively to the questions prompted by the Agent. Conversely, $\Lambda_B$ fails to define the relationships between the variables given by the formula and wrongly answers the questions prompted by the agent. By interpreting this evidence into the DDN, the INQPRO suggested a “partial mastery” level for $\Lambda_A$ while $\Lambda_B$ remains “non-mastery” for $H$, $\zeta$, and $K$.

Presumably that in time-slice $t_4$, INQPRO has captured meaningful interactions demonstrated by $\Lambda_A$ when $\Lambda_A$ revisits the Scenario GUI. These interactions include modifying the previously formulated hypothesis statement to a correct one and appropriately identifying variables. By taking the previous learning experiences (time-slice $t_3$) and current evidence into consideration, INQPRO has estimated a 45% mastery level for $H$, 55% for $\zeta$, and 48% of $K$ for $\Lambda_A$. $\Lambda_B$ who has also revisited the Scenario GUI conversely, did not manage to formulate the correct hypothesis and as such has resulted in the “non-mastery” level for $H$.

We shall present the evaluation of the proposed DDN model in the next section. A two-phase evaluation process was conducted aiming at investigating the matching accuracies and consistency of the proposed model in classifying learners’ $H$, $\zeta$, $K$, and $\Lambda$. We computed the matching accuracies based on a total of 76 learners and 6 domain experts.

**Evaluation of the Proposed DDN Model**

The evaluation phase of this research work aims at investigating the performance of our proposed DDN model for the INQPRO learning environment. At the first phase of evaluation $\Pi_1$, 76 first-year university learners participated in the evaluation while in the second phase of evaluation $\Pi_2$, there were a total of 31 first-year university learners involved. In both evaluation phases, learners were categorized into “High”, “Moderate”, and “Low” academic performance based on their O’level science or physics results. For both evaluation phases, learners participated in sessions that lasted between 80 to 100 minutes consisting of an introduction session to INQPRO, a pretest, a session with INQPRO and a posttest. Both the pretest and posttest consisted of 23 multiple choices questions with 9 questions targeted at assessing $H$, while 14 questions for $\zeta$. Due to the dependency between $H$, $\zeta$, and $K$, knowing the mastery levels of $H$ and $\zeta$ will help in predicting the mastery level of $K$. In this study, the pretest and posttest were validated by 2 domain experts who have years of experience in teaching Physics and Science. Learners who have scored higher than 75% for $H$, $\zeta$, and $K$ in both tests will be graded as “mastery” whereas those who scored less than 45% were regarded as “non-mastery”. Before interacting with INQPRO, learners were requested to elicit their own
mastery levels of $K$, $H$ and $\varsigma$ (hereafter ‘Pre-INQPRO self-ranking’). After the session with INQPRO, again, the learners were asked to rank their own acquisition level of $K$, $H$ and $\varsigma$ (hereafter ‘Post-INQPRO self-ranking’). A 3-rank scale (mastery, partial mastery, non-mastery) was given to the learners to assist them during the self-ranking process. In this study, however, the learners’ self ranking of mastery levels for $K$, $H$ and $\varsigma$ was not a primary index to calculate the performance of the proposed DDN model due to the fact that there were implicit and explicit factors which influenced the elicitation process.

In addition to the field tests administrated to the learners in $\Pi 1$ and $\Pi 2$, we consulted six domain experts to elicit their expectations on the proposed DDN model. At $\Pi 1$, the experts were presented with a 32 randomly selected learner interaction logs and videos, while at $\Pi 2$, there were a total of 14 randomly selected interaction logs were given to experts. The entire process lasted between 75 to 110 minutes.

At each phase of evaluation, we ran two experiments to obtain a more precise picture of the overall performance and stability of the proposed DDN model in assessing $H$, $\varsigma$, and $K$ of different categories of learners.

The first experiment aimed at examining how accurate and sensitive the proposed DDN model in classifying $H$, $\varsigma$, and $K$ for different categories of learners. To achieve this aim, we calculated the model’s matching accuracies by taking results obtained from the pretest and posttest of 107 learners as a reference. In this experiment, we computed the first set of the model’s matching accuracies by comparing the classifications given by the model with the classifications obtained from the pretest. In detail, those states with the highest probabilities for the `Static nodes` ($S^H$, $S^H$, and $S^H$ Figure 7) were compared to the classifications given by the pretest. At this stage of the experiment when the comparison is carried out, the DDN consisted of only nodes from the `Scenario GUI` mainly because the `Scenario` was the first GUI accessed by learners upon logging into the INQPRO learning environment. With the assumption that the actual learning has yet to occur for a particular learner within the `Scenario GUI`, the mastery levels of $H$, $\varsigma$, and $K$ should be reflected by the `Static nodes` through the states with the highest probabilities. Hence, we argue that such comparison approach is considered valid. Similarly, after interaction with INQPRO, the states of the `Static nodes` with the highest probabilities denote the ultimate mastery level of $H$, $\varsigma$, and $K$ for a particular learner. Again, we compare the classifications given by the model with the results obtained from the posttest to obtain the second set of the model’s matching accuracies. Having these two sets of matching accuracies, the overall average performance of the DDN model was computed and presented in terms of accuracies for $H$, $\varsigma$, and $K$.

The second experiment aimed at obtaining the matching accuracies for $H$, $\varsigma$, and $K$ from the experts’ point of view. In addition, this experiment also helped in identifying the appropriateness and suitability of decision alternatives (node $A_x$, $A_y$, which in this context were the pedagogical interventions $A$ generated by the `Agent`). The domain experts were consulted to elicit their predictions on the mastery levels of $H$, $\varsigma$, $K$, and $A$ at the GUI level. By studying the interaction logs for each GUI navigated by a learner, the experts verified the appropriateness of the pedagogical interventions ($A$) generated and estimated the mastery levels of $K$, $H$ and $\varsigma$. The matching accuracies of $K$, $H$, $\varsigma$, and $A$ were then calculated by counting the classifications given by a particular DDN model that matched those elicited by the experts. In addition to that, graphs (example see Figure 11 (a) and (b)) illustrating the evolving mastery levels of $K$, $H$ and $\varsigma$ were given to the experts aiming at investigating whether or not the patterns depicted fit their expectations. For each GUI, a $\checkmark$ was assigned to the matched classifications while an $\times$ was assigned to those misclassified.

Results

In this section, we shall firstly discuss the learners’ overall performance before and after interacting with INQPRO. Statistical significance tests were performed at both phase of evaluation to compare the individually matched improvements of the learners from pretest to posttest. Because the same pair of learners completed both a pretest and a posttest, a one-tailed paired difference experiment was performed to gauge the significant of the improvement.

We then present and analyze the accuracies of the proposed DDN model in classifying the mastery levels of $K$, $H$ and $\varsigma$ for different categories of learners. We shall begin with the first phase of evaluation $\Pi 1$ and subsequently with the second phase of evaluation $\Pi 2$. The results of the experiments at each evaluation phase shall be addressed and
discussed in detail. In this study, we use the pretest and posttest results as benchmark for the computation of matching accuracies. The self ranking, however, is not a benchmark to how accurate the proposed DDN performs due to the learners’ implicit and explicit factors that might influence the learners in self ranking process.

First Phase of Evaluation (Π1)

Figure 12 depicts an overview performance demonstrated by the learners at the first phase of evaluation. The mean score for pretest was 16.3 while the mean score for posttest was 36.4. The standard deviation for pretest was relatively high (10.9) compare to the posttest with 6.7. It was found that the \( p\)-value was 1.2E-07 and thus concluded that the t-test confirmed that the improvement from pretest to posttest was significant at a \( p<0.05 \) level.

![Figure 12. Mean pretest and posttest scores for 46 learners](image)

Experiment 1: Model’s matching accuracies using pretest and posttest as reference

Table 1 shows the results of the proposed DDN model in classifying different categories of learners. As shown in Table 1, the model demonstrated an overall matching accuracies of higher than 50% for the pretest, posttest, pre-INQPRO self ranking, and post-INQPRO self ranking. The high matching accuracies given at the pretest suggested that the proposed DDN performed well in classifying learners’ initial acquisition of \( H \) (97.8%) and the understanding of overall scientific inquiry skills \( K \) (93.5%). The variation of matching accuracies for \( K \) and \( H \) within 10% with respect to the self-ranking columns has also suggested the model’s consistency in modeling learners’ scientific inquiry skills. In this experiment, the reliability of self ranking scores is not accountable due to personal traits. Thus, taking them away would give the overall matching accuracy for \( K=85.9\% \) while \( H=86.9\% \). Such promising results indicate that the proposed DDN model was able to perform well in classifying learners’ \( H \) and \( K \).

Table 1 has, however, depicted a relatively low matching accuracy for \( \varsigma \) with respect to the pretest (52.2%). Such low accuracy was mainly due to the misclassification of learners into “partial-mastery” by the DDN model whereas the pretest categorized them as “non-mastery”. Through the interview sessions with randomly selected learners, we found out that the learners learned from pretest on \( H \) and \( \varsigma \) and such learning effect had subsequently become the prior knowledge as they interacted with INQPRO. As a result, misclassification happened and a low accuracy for \( \varsigma \) at the pretest is shown (Table 1, \( \varsigma=52.2\% \)). Such phenomenon was also reflected by the low accuracy depicted at the pre-INQPRO self ranking for \( \varsigma \) (47.8%). These learners had ranked themselves as “non-mastery” before the pretest.

The results shown in Table 1 also highlights that the proposed DDN model performed equivalently well in classifying learners from the moderate and low category. Taking the moderate group as an example, the proposed
DDN model successfully classified 91.7% for $K$, 83.0% for $H$, and 83.3% for $\varsigma$ with respect to the Posttest column. The ability to classify learners from different categories is a crucial part for a good classifier.

Table 1. The model’s matching accuracies with respect to results obtained from pretest, posttest, and self-ranking during the first phase of evaluation $\Pi_1$ (46 learners)

<table>
<thead>
<tr>
<th>Scientific Inquiry Skill</th>
<th>Category of learners</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Pre-INQPRO</th>
<th>Post-INQPRO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Advance ($n=13$)</td>
<td>10(90.9)</td>
<td>9(81.8)</td>
<td>2(18.2)</td>
<td>7(63.6)</td>
</tr>
<tr>
<td></td>
<td>Moderate ($n=20$)</td>
<td>11(91.7)</td>
<td>11(91.7)</td>
<td>7(58.3)</td>
<td>7(58.3)</td>
</tr>
<tr>
<td></td>
<td>Low ($n=43$)</td>
<td>22(95.7)</td>
<td>16(69.6)</td>
<td>15(65.2)</td>
<td>9(39.1)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>43(93.5)</td>
<td>36(78.3)</td>
<td>24(52.2)</td>
<td>23(50.0)</td>
</tr>
<tr>
<td>$H$</td>
<td>Advance ($n=13$)</td>
<td>11(100.0)</td>
<td>9(82.0)</td>
<td>2(18.2)</td>
<td>7(63.6)</td>
</tr>
<tr>
<td></td>
<td>Moderate ($n=20$)</td>
<td>11(91.7)</td>
<td>10(83.0)</td>
<td>7(58.3)</td>
<td>6(50.0)</td>
</tr>
<tr>
<td></td>
<td>Low ($n=43$)</td>
<td>23(100.0)</td>
<td>16(69.6)</td>
<td>14(60.9)</td>
<td>14(60.9)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>45(97.8)</td>
<td>35(76.1)</td>
<td>23(50.0)</td>
<td>27(58.7)</td>
</tr>
<tr>
<td>$\varsigma$</td>
<td>Advance ($n=13$)</td>
<td>8(82.7)</td>
<td>10(90.9)</td>
<td>7(63.6)</td>
<td>9(81.8)</td>
</tr>
<tr>
<td></td>
<td>Moderate ($n=20$)</td>
<td>5(41.7)</td>
<td>10(83.3)</td>
<td>5(41.7)</td>
<td>6(50.0)</td>
</tr>
<tr>
<td></td>
<td>Low ($n=43$)</td>
<td>11(47.8)</td>
<td>11(47.8)</td>
<td>10(43.5)</td>
<td>18(78.3)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>24(52.2)</td>
<td>31(67.4)</td>
<td>22(47.8)</td>
<td>33(71.7)</td>
</tr>
</tbody>
</table>

As a conclusion, the proposed DDN model demonstrated its consistency in classifying different categories of learners with an average accuracies of 85.9% for $K$, 86.9% for $H$ while 59.8% for $\varsigma$. In the next subsection, we shall address the second evaluation objective by investigating the behaviours and matching accuracies of the proposed DDN model at GUI level elicited from the domain experts. This step is crucial as it helps in verifying the consistency and appropriateness of the proposed DDN model before conclusion on the performance is made.

Experiment 2: Model’s matching accuracies through domain experts

The importance of having domain experts participating in this study is two-fold: (i) to verify the classifications depicted by the model at each individual GUI given a set of learner interactions as evidences; (ii) to examine the appropriateness and suitability of pedagogical interventions generated by the $Agent$ at each GUI. Such elicitation process has allowed us to study how similar the pedagogical interventions generated by the model as compared to those from the experts.

Table 3. The model’s matching accuracies with respect to domain experts

<table>
<thead>
<tr>
<th>% matched classification</th>
<th>Advance ($n=4$)</th>
<th>Moderate ($n=12$)</th>
<th>Weak ($n=2$)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$</td>
<td>91.7</td>
<td>84.4</td>
<td>91.7</td>
<td>89.3</td>
</tr>
<tr>
<td>$H$</td>
<td>97.2</td>
<td>80.5</td>
<td>95.8</td>
<td>91.2</td>
</tr>
<tr>
<td>$\varsigma$</td>
<td>91.3</td>
<td>80.3</td>
<td>87.5</td>
<td>86.4</td>
</tr>
<tr>
<td>$A$</td>
<td>98.4</td>
<td>93.4</td>
<td>100</td>
<td>97.3</td>
</tr>
</tbody>
</table>

Table 3 shows the matching accuracies of the proposed DDN model as elicited by the domain experts. The model has depicted equivalent promising results for the three categories of learners with respect to $K$ (89.3%), $H$ (91.2%), $\varsigma$ (86.4%), and $A$ (97.3%). With the matching accuracies exceeding 80%, we can ascertain that the proposed DDN model had demonstrated its adaptability in classifying $K$, $H$, and $\varsigma$ that evolved across time. In addition, the high matching accuracy depicted for $A$ (96.1%) in Table 3 also indicated that the experts had somehow agreed with the pedagogical interventions generated by the $Agent$. During the interview sessions, the experts had commented that it was a difficult process to justify whether or not a particular learner had achieved mastery levels of $\varsigma$ and $H$ until the
correct hypothesis statement was constructed and appropriate variables were selected. Due to the constraint, low matching accuracies were recorded for H (80.5%) and ζ (80.3%) for the moderate group. When the domain experts were asked regarding the effects of interacting with the Hypothesis Visualization GUI to the acquisition of K, H, and ζ, they strongly agree that active interactions could help improve learner’s acquisition of scientific inquiry skills. However, again, the domain experts require solid evidence before concluding the mastery level of a learner.

**Evaluation Phase Two (Π2)**

Figure 13 depicts an overview performance demonstrated by the learners at the second phase of evaluation. The mean scores for both pretest and posttest at the second phase of evaluation were relatively low compared to the first phase of evaluation. However for both evaluation phases, the standard deviations were similar for both pretest and posttest. When t-test was performed, it was found that the *p*-value was 7.8E-05 and thus concluded that the overall improvement demonstrated by the learners from pretest to posttest was significant at a *p*<0.05 level.

![Figure 13. Mean pretest and posttest scores for 31 learners](image)

**Experiment 1: Model’s matching accuracies using pretest and posttest as reference**

A total of 31 learners participated at Π2. Again, the matching accuracies of the model depicted in Table 4 were calculated by comparing the classifications given by the model with those obtained from pretest, posttest, and self-ranking scores.

Table 4 shows the matching accuracies of three different categories of learners. Without taking the matching accuracies of self-ranking into consideration, the average matching accuracies for both H and ζ reached 85.5%. As shown in Table 4, the overall matching accuracies for ζ was again relatively low (pretest=64.5%; pretest self ranking = 58.1%) as compared to the accuracies depicted by H and K. The reason is identical to that we have analyzed and concluded in Π1 where the “learning” effect experienced by the learners during the pretest had somehow helped them recall scientific inquiry skills. When such “learning” effect was transferred and incorporated into the DDN model, the mismatch of classifications occur.

Another point worth highlighting in this evaluation is the relatively low matching accuracy for H (74.2%) shown in Table 1. There were two major reasons for this issue. Firstly, from the observation as the learners worked individually with INQPRO, we found out that there were learners who did not perform meaningful interactions. These learners did not interact with the Agent, and had not performed enough interactions such as drag-and-drop the different masses at the Hypothesis Visualization GUI. Due to the lack of interactions, INQPRO would treat the learners as “unable to interact successfully” and when these evidences are fed into the network, the resulting
categories were mostly likely to be either “partial mastery” or “non mastery”. The second reason responsible for the low accuracy was due to the uncompleted posttest. At this evaluation phase Π2, there were two learners who did not complete the posttest. As a result, the mismatch of classifications occurred. However, if these two reasons were not taken into consideration when it came to calculating the matching accuracies, the matching accuracy for H for the posttest section would then be 88.5%.

Table 4. The model’s matching accuracies with respect to the results obtained from pre-test, post-test, and self-ranking during the second phase of evaluation Π2 (31 learners)

<table>
<thead>
<tr>
<th>Scientific Inquiry Skill</th>
<th>Category of learners</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Pretest Self-Ranking</th>
<th>Posttest Self-Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Advance (n=8)</td>
<td>7(87.5)</td>
<td>7(87.8)</td>
<td>5(62.5)</td>
<td>4(50.0)</td>
</tr>
<tr>
<td></td>
<td>Moderate (n=14)</td>
<td>11(78.6)</td>
<td>13(92.9)</td>
<td>7(50.0)</td>
<td>7(50.0)</td>
</tr>
<tr>
<td></td>
<td>Low (n=9)</td>
<td>8(88.9)</td>
<td>7(77.8)</td>
<td>7(77.8)</td>
<td>4(44.4)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>26(83.9)</td>
<td>27(87.1)</td>
<td>19(61.3)</td>
<td>15(48.4)</td>
</tr>
<tr>
<td>H</td>
<td>Advance (n=8)</td>
<td>7(87.5)</td>
<td>6(75.0)</td>
<td>6(75.0)</td>
<td>5(62.5)</td>
</tr>
<tr>
<td></td>
<td>Moderate (n=14)</td>
<td>14(100.0)</td>
<td>12(86.0)</td>
<td>5(35.7)</td>
<td>7(50.0)</td>
</tr>
<tr>
<td></td>
<td>Low (n=9)</td>
<td>9(100.0)</td>
<td>5(55.6)</td>
<td>6(66.7)</td>
<td>5(55.6)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>30(96.8)</td>
<td>23(74.2)</td>
<td>17(54.8)</td>
<td>17(54.8)</td>
</tr>
<tr>
<td>ζ</td>
<td>Advance (n=8)</td>
<td>5(62.5)</td>
<td>7(87.5)</td>
<td>3(37.5)</td>
<td>8(100.0)</td>
</tr>
<tr>
<td></td>
<td>Moderate (n=14)</td>
<td>8(57.1)</td>
<td>13(92.9)</td>
<td>8(57.1)</td>
<td>13(92.9)</td>
</tr>
<tr>
<td></td>
<td>Low (n=9)</td>
<td>7(77.8)</td>
<td>5(55.6)</td>
<td>7(77.8)</td>
<td>8(88.9)</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>20(64.5)</td>
<td>25(80.7)</td>
<td>18(58.1)</td>
<td>29(93.6)</td>
</tr>
</tbody>
</table>

Experiment 2: Model’s matching accuracies through domain experts

Table 5 shows the averages for the matching accuracies for K (93.3%), H (83.1%), ζ (82.9%), and A (85.4%). With all the average accuracies exceeding 80%, once again, the proposed DDN model was highly rated for its appropriateness in modeling learners’ scientific inquiry skills and adaptability in providing pedagogical interventions for different categories of learners. In this experiment, again, the experts rated ζ (82.9%) and H (83.1%) with a relatively low accuracy due to the similar reasons given at Π1. Compared to Π1, Π2 has more learners who did not navigate back to the Scenario GUI to correct the hypothesis statements. In addition, with only one learner categorized under the weak category, this again had contributed to the relatively low accuracy recorded for H and ζ (Table 5).

Table 5. The model’s matching accuracies with respect to domain experts

<table>
<thead>
<tr>
<th></th>
<th>Advance (n=4)</th>
<th>Moderate (n=9)</th>
<th>Weak (n=1)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>93.2</td>
<td>95.7</td>
<td>90.9</td>
<td>93.3</td>
</tr>
<tr>
<td>H</td>
<td>85.5</td>
<td>86.4</td>
<td>77.3</td>
<td>83.1</td>
</tr>
<tr>
<td>ζ</td>
<td>84.9</td>
<td>91.0</td>
<td>72.7</td>
<td>82.9</td>
</tr>
<tr>
<td>A</td>
<td>92.4</td>
<td>90.5</td>
<td>73.3</td>
<td>85.4</td>
</tr>
</tbody>
</table>

Conclusion and Future Directions

The research work presented in this article aimed at proposing a methodological approach for modeling and intervening under uncertainty within the INQPRO learning environment. In this study, we scope the scientific inquiry skills to Hypothesis Formulation H and Variables Identification ζ. During the model construction phase, we introduced K to represent the overall understanding of H and ζ because of the causal relationship between them. In this article, we highlighted how our proposed methodological approach has complemented and contributed to the
existing research work in science education and intelligent tutoring systems particularly on those leveraging decision-theoretic approach. From science education perspective, assessing a learner’s scientific inquiry skills had always been a challenging task (de Jong, 2006). The pencil-and-paper approach is often deployed as an assessment method. Such method, although commonly implemented, has its limitations as it only allows the assessment of the final but not the evolving mastery levels of scientific inquiry skills. In this light, one of the aims of the proposed methodological approach was to provide a mechanism to explicitly visualize the evolving mastery level as the learners interact with a particular computer based learning environment. From the ITS perspective, most of the existing decision-theoretic tutoring systems rely on a DDN or DBN that is constructed from a repeated static BN or DN. This approach, however, is not always practically sound to all the learning environments. The INQPRO deployed in this study for instance, consists of six different GUls with different DNs correspond to them.

Overcoming these challenges is not a trivial task. Various considerations in constructing the DDN need to be well taken care of. In this study, the proposed novel methodological approach consisted of (i) a dynamically generates a DDN from individual DNs at run-time basis. Once a particular GUI was navigated by a learner, the corresponding DN is appended to the newly created DDN; (ii) inclusion of the static nodes in the newly created DDN at run-time basis. Once the DDN was generated, the static nodes were added on top of the nodes aggregated from separated DN.

In this article, we have presented two evaluation phases, Π1 and Π2, that aimed at investigating the behaviours and performance of the proposed DDN model. A total of 46 learners participated at Π1 while 31 learners participated at Π2. There were a total of 6 domain experts participating at both evaluation phases. At each evaluation phase, we conducted 2 experiments. The first experiment relied on the pretest, posttest, and self ranking scores as a benchmark for calculating the matching accuracies while the second experiment depended on the classification results given by the domain experts as a reference. For both evaluation phases, we found out that the matching accuracy for ς with respect to the pretest was relatively low. This was mainly due to the misclassification of learners into “partial-mastery” by the DDN model whereas the pretest categorized them as “non-mastery”. The interview sessions with randomly selected learners from both phases revealed that the learners learned about ς and H from the pretest, and such learning effect became the prior knowledge that equipped them with the necessity to interact with INQPRO. A slightly low matching accuracy for ς and H at the posttest for both evaluation phases is mainly because of the unexpected system faults such as the runtime errors, and program termination. Apart from that, the relatively low matching accuracy for H at the posttest was also due to the incompleteness of meaningful interactions with INQPRO demonstrated by the learners. Due to lack of evidence, the system interpreted these learners as “non-mastery”. These learners, although they had not interacted successfully with INQPRO, had scored well at the posttest. Due to the mismatched of such circumstances, the accuracy of H at the posttest was relatively low.

Employing DDN approach in this study, however, created a contradiction to the conventional assessment method employed by domain experts. Employing DDN in this study has allowed the modeling of the human interventions during the model

The current research work can be further enhanced and tackled from two perspectives: (i) the structural aspect of the proposed DDN model; and (ii) the parameters or beliefs of the nodes in each DN. The current DDN model is having d-separation connection for the static nodes (H \rightarrow K, and ς \rightarrow K). The arc stretching from ς to H indicates a causal relationship between these two nodes, which suggests that increasing the probability of ς will also increase the probability of H. The proposed approach has, however, resulted in the tendency of the proposed model to over interpret of evidence upon H due to the propagation of probabilities from ς to H and from K to H.

One of the great advantages of employing the proposed approach is that it can be directly deployed without the need to train the model beforehand. In this study, the parameters setting of the proposed DDN model can be incorporated from the knowledge elicited from the domain experts. However, such a method does not guarantee an optimal solution to model and intervene under uncertainty within the INQPRO learning environment. Thus, our immediate next step is to apply the Bayesian machine learning approach. We hope that by applying the machine learning approach to learn both the structure and the beliefs of the DNs, the human interventions during the model
construction can be minimized and at the same time obtaining an optimum solution to the modeling and intervening issues can be obtained.

References


A Practical Computer Adaptive Testing Model for Small-Scale Scenarios

Yu-Hui Tao
Department of Information Management, National University of Kaohsiung, Taiwan, R.O.C. // ytao@nuk.edu.tw // Tel: +886-7-5919220 // Fax: 886-7-5919328

Yu-Lung Wu
Department of Information Management, I-Shou University // wuyulung@isu.edu.tw

Hsin-Yi Chang
Corel Corporation, Taiwan, R.O.C. // jasminechang@ulead.com.tw

ABSTRACT

Computer adaptive testing (CAT) is theoretically sound and efficient, and is commonly seen in larger testing programs. It is, however, rarely seen in a smaller-scale scenario, such as in classrooms or business daily routines, because of the complexity of most adopted Item Response Theory (IRT) models. While the Sequential Probability Ratio Test (SPRT) model is less complicated, it only provides the examinee’s mastery result. We thus propose an SPRT-based adaptive testing approach that is simpler to implement while still being able to approximate IRT scores in semantic or rank levels for flexible assessment needs. An English adaptive testing prototype is implemented and benchmarked to the Test of English as a Foreign Language (TOEFL) testing results. Generally, this research empirically demonstrates the validity of the proposed SPRT-based testing approach as well as the technical feasibility for teachers and business organizations to really take advantage of CAT in their daily routines.

Keywords
Computer adaptive testing, Item response theory, Sequential probability ratio test, TOEFL

Introduction

Traditionally, testing for evaluating knowledge, skills, abilities, and other characteristics (KSAO’s) has been done in a paper-and-pencil scenario. However, the development of information technology (IT) in the last two decades has made computer-based testing (CBT) feasible in both educational research and practice (Bunderson et al., 1989). Furthermore, today’s e-learning technology enables organizations to start adopting online instructions as well as online testing. The evolving technologies have thus moved the traditional pencil-and-paper testing toward a computer-based, or even a computer adaptive testing (CAT) scenario.

In theory, CAT can dramatically reduce the testing time while maintaining the quality of measurement as compared to the fixed-item type of tests in either pencil-and-paper or CBT format (Wise & Kingsbury, 2000). Thus, it has been researched and applied extensively in larger educational institutes, certified or licensed centers (Olson, 1990; ETS, 2001; Taiwan Education Testing Center, 2007). However, CAT is not used by either classroom teachers who make up and administrate their own tests (Frick, 1992) or by business organizations in their daily KSAO’s routines. One major cause for this situation is that the most adopted CAT model - Item Response Theory (IRT) - is too rigorous to implement and maintain. Wise and Kingsbury (2000) listed item pools, test administration, test security, and examinee issues as the four general areas of practical issues in developing and maintaining IRT-based CAT programs. In particular, the adopting issues mostly fall into the item pools area, which includes pool size and control, dimensionality of an item pool, response models, item removal and revision, addition of items to the item pool, maintenance of scale consistency, and the use of multiple item pools. Since the rigorous IRT requires a large number of examinees ranging from 200 to 1000 for estimating item parameters and special expertise in item-pool maintenance, IRT is only possible in educational institutes or professional testing centers (Frick, 1992).

The Sequential Probability Ratio Test (SPRT) model is another CAT model that is less adopted because it only provides the examinee’s mastery result and lack of the assessment flexibility of the IRT score. Nevertheless, the original SPRT waives the maintenance requirements for the pre-test that involves a large number of examinees (Frick, 1990). This characteristic of SPRT also suffers issues in variability in item difficulty, discrimination, or chances of guessing. An empirical study of Frick (1990) indicated that SPRT is a fairly robust model for mastery
decisions, especially under smaller Type I and II decision error rates such as 0.025. Moreover, although parameter estimation pre-test and calibration on item pool may be preferred, IRT still suffers from accuracy or validity issues (Frick, 1990; Huff & Sireci, 2001; Welch & Frick, 1993). From the above perspectives, SPRT seems to be a practical alternative to the CAT application for school teachers and business organizations.

We propose an SPRT-based CAT approach that inherits the maintenance-free item pool of SPRT strength, and approximates the grade classification of IRT spirit. In addition, to show the validity of our proposed approach, the criterion validity (Zikmund, 1997) method is adopted by comparing an English CAT prototype system based on the proposed approach with the “Test of English as a Foreign Language (TOEFL)” standard. Criterion validity was chosen because the potential source of construct-irrelevant variance originating from one’s unfamiliarity with computers had been studied and concluded to be negligible (Taylor et al., 1999). Technically speaking, the criterion validity basically answers questions like “Does my measure correlate with other measure of the same construct?” (Zikmund, 1997). In other words, if the proposed SPRT-based CAT approach can distinguish the English abilities among examinees just as TOEFL does, it can be claimed to establish the criterion validity against the TOEFL. Accordingly, our benchmark test against TOEFL can be a leading indicator for the empirical validity of the proposed SPRT-based approach.

Background

CAT differs from the traditional pencil-and-paper or regular CBT testing in that an evaluation is done with a possible minimum number of questions adaptive to the ability of the examinee (Welch & Frick, 1993). Since both IRT and SPRT are the primary models in adaptive testing, and TOEFL is the benchmarking standard in our experiment, they are briefly introduced below.

IRT

Lord (1980) invented IRT in the early 1950s, which utilized probability to explain the relationship between the examinee’s ability and the question item response. Specifically, a mathematical model, called Item Characteristic Function (ICF) that derives a continuous increasing curve for predicting the ability and test performance, was developed to infer the examinee’s ability or potential quality.

The ICF model can be classified into different variations based on the number of parameters adopted within the mathematical model. The three often-used models include single-parameter, two-parameter, and three-parameter models, which is well summarized by Yu (1992) with original references:

Single-parameter model

\[ P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}} , \]  

\[ (1) \]

Two-parameter model

\[ P_i(\theta) = \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} , \]  

\[ (2) \]

Three-parameter model

\[ P_i(\theta) = c_i + (1-c_i) \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} , \]  

\[ (3) \]

where \( D=1.702 \) 
\( e \): natural Log, 2.71828 
\( i \): item number; \( i=1,2,3,...,N \), and \( N \) is the total number of items 
\( \theta \): examinee’s ability 
\( a_i \): discrimination parameter of item \( i \) 
\( b_i \): difficulty parameter of item \( i \) 
\( c_i \): guess parameter of item \( i \)
IRT has a few basic assumptions, including unidimensionality, local independence, non-speeded test, and know-
correct assumption, which need to be sustained before the model can be used to analyze the data (Ackerman, 1989).
In IRT theory, an item is treated as the basic unit for measuring the examinee’s ability through the parameters. The
execution follows a simple principle: if an examinee responds correctly to an item, then the next item will be one
level up in terms of difficulty, and vice versa. For each item responded to, the examinee’s estimated ability will be
reevaluated, and then the appropriate level of the next item is given. The process is repeated until a pre-set reliability
level or stopping rule is triggered (Huff & Sireci, 2001).

IRT has been widely used in CAT studies, such as those focusing on ability testing (Si, and Schumacker, 2004), CAT
systems (Chou, 2000; Ho & Yen, 2005; Guzmán & Conejo, 2005) and e-learning (Yang et al., 2005; Hatzilygeroudis
et al., 2006; Wang, 2006).

SPRT
Wald (1947) originally developed SPRT and Ferguson (1967) later adopted SPRT in education testing for the pass-
or-fail decision. Although SPRT is easier to be implemented and thus adopted by Wang & Chuang (2002) for school
teachers, it is not commonly used in education (Welch & Frick, 1993) but it is used in other domains such as
Engineering (Thanh Ngoc Bui et al., 2004; Das et al., 2005) and Science (Tartakovsky et al., 2003; Jarman et al.,
2004).

In the standard SPRT execution, items were randomly selected, and the sequential probability ratio was calculated
based on the item response. Related definitions of terminologies found in the study of Frick (1990) are as follows:

\[
LBM (\text{Lower Bound Mastery}) = \frac{(1-\beta)}{\alpha} , \tag{4}
\]

\[
UBN (\text{Upper Bound Nonmastery}) = \frac{\beta}{(1-\alpha)} , \tag{5}
\]

Probability of correct item response
\[
PR = \frac{P_m^s (1 - P_m)^f}{P_{nm}^s (1 - P_{nm})^f} \tag{6}
\]

where \(P_m\) = Probability of mastery with correct item response
\(P_{nm}\) = Probability of non-mastery with correct item response
\(s\) = number of correct item responses out of the total number of items responses so far
\(f\) = number of wrong item responses out of the total number of items responses so far
\(\alpha\) = Type I error, judging mastery, but in fact non-mastery
\(\beta\) = Type II error, judging non-mastery, but in fact mastery.

When an examinee finishes an item and the system-calculated PR is greater than or equal to LBM, the result of the
examinee’s test is judged to be “mastery” and the test is terminated. If the result is undetermined or “non-mastery”
with UBN < PR < LBM, then the test goes on with a new randomly selected item. Otherwise, if PR < UBN, the
examinee is judged to be “non-mastery” (Frick, 1989). Although SPRT does not involve complicated mathematical
formulas, there are still two basic assumptions: A test item is randomly selected from the item bank and cannot be
repeated, and it has local independence such as the one in the IRT.

Two requirements may prevent an instructor from developing the item pool. First, Frick (1992) reported that IRT
requires 200 to 1000 examinees to adjust the parameters, which consumes more time and human resources than
SPRT in preparation. Second, IRT involves complex mathematical formulas. As compared to the complexity of IRT
execution, SPRT is a simpler model to implement. However, its potential drawbacks are as follows: not considering
item difficulty, discrimination, and guess degree. Therefore, Frick mixed these two theories and developed an expert
system-based SPRT (EXSPRT). The application of EXSPRT is similar to SPRT, but the former weighted differently
on the items in the item pool. Therefore, EXSPRT requires only 50 examinees to adjust the item bank (Frick, 1992).
A comparison among IRT, SPRT, and EXSPRT is summarized in Table 1.
Table 1. Comparison between adaptive models

<table>
<thead>
<tr>
<th>Adaptive model</th>
<th>Examinee size for parameter estimation</th>
<th>Item selection</th>
<th>Difficulty of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRT</td>
<td>200-1000 people</td>
<td>Item response</td>
<td>Difficult</td>
</tr>
<tr>
<td>SPRT</td>
<td>None</td>
<td>Random</td>
<td>Easy</td>
</tr>
<tr>
<td>EXSPRT</td>
<td>50 people</td>
<td>Random in EXSPRT-R or Item response in EXSPRT-I</td>
<td>More difficult than SPRT</td>
</tr>
</tbody>
</table>

We would like to emphasize that even though SPRT has been used inappropriately, such as in items varying widely in difficulty levels, and/or discriminating power is reaching a master or a non-master decision before a representative sample of items has been administered to an examinee, the SPRT decision are highly accurate when prior error rates ($\alpha$ and $\beta$) are kept at their minimum, as low as 0.025 (Frick, 1990).

TOEFL

TOEFL is an English proficiency proof for applying in universities or graduates schools in Canada and the United States, it is provided by Educational Testing Service (ETS) for the non-native English speakers (ETS, 2007). Taiwan’s TOEFL has implemented the CAT since October 2000 where each examinee is equipped with a computer and a headphone in the test site. The TOEFL CAT test only allows the examinee to see one item at a time on the screen. When an examinee presses the enter key, the computer evaluates the response and selects the next item with an appropriate difficulty. The examinee can neither skip any item nor go back to previous items.

The proposed SPRT-based CAT approach

The main objective of this research is to propose a feasible solution for CAT to be used in typical classrooms or business organizations. The idea of the proposed approach is that it would be SPRT-based and would adopt the IRT spirit of the non-mastery result. As a base, SPRT provides its simplicity in the item pool which consists only of items with equal levels of difficulty. The IRT spirit of the non-mastery result is achieved by expanding the SPRT-based item pool into an N-tier item pool. In other words, each tier of the item pool contains only the same difficult level of items, so that an examinee’s ability can be identified to be up to one of the N levels. The rationale behind this design is that the domain experts/teachers in-charge can easily maintain the N-tier item pool based only on his/her professional expertise. No additional expertise in testing theory and parameter-estimation process is required, as in EXSPRT or IRT.

Figure 1. IRT CAT model
In order to illustrate this idea more clearly, Figure 1 shows a simplified diagram of a typical IRT-based CAT, while Figure 2 shows the proposed SPRT-based CAT. Two major differences can be easily identified from these two diagrams: the IRT includes an additional parameter-estimation ($a_i$, $b_i$, and $c_i$) pre-test task for the formulas (1)-(3) to work, while the SPRT-based CAT has the item pool divided into N tiers by splitting various difficulty levels of IRT items into cohesive groups.

![Figure 2. SPRT-based CAT approach](image)

In creditable certified or license centers, it is even required that the SPRT item pool be rigorously verified against the three-parameter logistic model (PLM) IRT as suggested by Kalohn & Spray (1999). However, in order for this SPRT-based CAT approach to be easily adopted in daily routines, we assume that the tedious parameter-estimation task may be omitted as long as certain validity is retained in it. Here the certain validity means that SPRT is robust under small prior error rates ($\alpha$ and $\beta$) as seen in Section 2.2, which is probably acceptable to the teachers or managers in small-scale scenarios compared to the large testing centers. Therefore, a question remains to be answered: how does the outcome of the proposed approach compare to the IRT score? Since each tier of the item pool represents a different scale of ability, an examinee goes up and down a certain ability level and through the SPRT item pool from a certain ability level before finally stopping at the first tier judged to be non-mastery or undetermined. Then the examinee obtains a semantic label of ability, such as medium or proficient, or an ability level, such as level 1 or level 3. Note that the ability level still requires further interpretation, such as 1 indicates proficient or novice, while the semantic label is already an interpreted ability-level in human language.

To see how the SRPT-based CAT approach works, an integrated process for the CAT example is illustrated in Figure 3. The upper half of Figure 3 depicts a single SPRT process cycle, while the lower half represents the N-tier SPRT algorithm that determines how to repeat the single SPRT process cycle. In our CAT scenario, three objectives of items include Listening, Grammar, and Reading, with each having six difficulty levels that are set up in the first two actions in the SPRT base.

For the most complex requirement satisfying different perspectives of English ability, the appropriate ability levels of all three objectives for an examinee need to be identified. If an examinee starts with Grammar, the default tier is level 3, and an item is randomly selected from the third tier item pool. When an item response is evaluated to be $\text{PR} \geq \text{LBM}$, then level 3 would be judged to be “mastery” and level 4 is triggered. If $\text{PR} \leq \text{UBN}$, then level 3 is judged to be “non-mastery” and level 2 is triggered. Otherwise, $\text{UBN} \leq \text{PR} \leq \text{LBM}$ means that level 3 is undecided. Then the next item will be randomly selected again for the examinee to answer if the stopping condition has not yet been met. When the ability level of Grammar is determined, then either Listening or Reading is triggered. This complex scenario does imply a tradeoff in this SPRT-based CAT approach for flexible assessment needs. That is, it takes longer than the IRT CAT to decide the examinee’s ability due to the N-tier structure that repeats the SPRT test up to N times.
The simplest scenario would be to determine whether or not an examinee would achieve “mastery” on a certain level in one of the Grammar, Listening, or Reading objectives. The process is similar to what has been described in the most complex scenario, except that no other level will be triggered when the assigned mastery level is decided.

Prototype system

A Web-based English CAT prototype system was constructed to validate the feasibility and correctness of the proposed approach as shown in Figure 4. Four basic modules are included – administrative users, maintain item pool, setup test, and take test.

The “administrative user” module contains basic data for human-resource (HR) staff/teachers who manage the test process, employees or applicants/students who take the test, and the domain experts/teachers who maintain the test item pools. The “maintain item pool” module provides maintenance functions for adding, editing, deleting, or previewing test items, which supports an easy way to maintain the quality of the item pool without CAT or IT expertise. The “setup test” module setup refers to online adaptive tests for evaluating the English ability of examinees, while the “take test” module accepts examinees logging in to precede a pre-set personalized test session. All the testing history and records are automatically stored for query purposes on a needed basis.
The prototype system was developed on a Windows server running a Java-based Web server Resin and a Microsoft ACCESS database. The Java Server Page (JSP) Web programming language was used to implement the logic of interactive programming. The purpose of the interactive screens is to illustrate how easily domain experts, HR staff, and teachers can maintain the item pool or administrate a test, and employees or job applicants/students can take the test. Figure 5 shows how an item was added to the pool on a screen with Item Level (1-6 difficulty), Item Type (Grammar, Reading and Listening), question, and the answer options with a radio-box checked as the correct answer. Figure 6 then illustrates how to browse the Grammar item pool and select an item for modifications. Similar screens of interactive style are provided for deleting and previewing an item, which are not shown here.
Figure 6. Browsing the grammar item pool

Figure 7. The objective of Listening item testing screen
Figure 6 shows an examinee answering a Listening item in difficulty level 3. The item representation is the same in both Listening and Grammar objectives, except that the question and maybe the options in Listening are those of the voice data type. Therefore, the examinee needs to press the yellow speaker icon to hear the question or options, and once heard, the little icon is changed to a disabled status with an extra red cross on top of the yellow speaker as shown in Figure 7.

Reading items, by their nature, form a group of cohesive items corresponding to an article. Therefore, as shown in Figure 8, an article is displayed on top of the test item, and the cohesive group items are displayed one at a time for the examinee to answer. The question, as usual, has to be answered and cannot be skipped, but the article will remain until all cohesive items related to that article are finished by the examinee.

Principally, the examinee navigates the prototype system as depicted in Figure 3, where some of the actual screen shots are presented in Figures 5-7. The teacher who is in charge of administrating the test sets up the navigation and terminating conditions in advance. In a fixed-item test, an examinee may be able to browse through all or parts of the questions repeatedly before finally submitting the answers. When a student takes a pre-assigned adaptive test, the navigation route is automatically guided by the system with the required-to-know information on screens since there are many concurrent background decisions as illustrated in Figure 3. As demonstrated in Figures 5-7, all users can interact with the CAT system easily relating to their tasks. Therefore, there will be no need to worry about IT implementation issues or CAT testing expertise, since this SPRT-based CAT approach can be as feasible as the prototype demonstrated in this section.

Experiment design and settings

With an initial prototype system, a pre-test was first conducted to four subjects for assessing the appropriateness of the system and interface designs. Consequently, the formal experiment was conducted only after the opinions and
suggestions collected from the pre-test are adjusted in the prototype system. Since the pre-test followed the exact same procedure as the formal experiment, only the design of the formal experiment is summarized below.

The goal of this experiment is to evaluate the effectiveness of the proposed approach by benchmarking our English CAT prototype system with the well-established TOEFL test, which belongs to the predictive validity method in criterion validity (Zikmund, 1997). Due to no available sample name list, this study located the first 60 accessible working professionals with TOEFL scores in Taiwan. Each experiment session started with a 20-minute brief introduction of this experiment and a warm-up practice of the English CAT prototype system. Then an official online CAT test session of 80 minutes on English Grammar, Listening, and Reading was given to the examinee, followed by a 20-minute interview.

The data analysis proceeded in two steps. First, the subjects were divided into groups based on their TOEFL scores, which were double-validated by the independent Mann-Whitney test for their significant differences. Then with the same groups, the average ability levels collected from our experiment were calculated, and the Mann-Whitney test was again applied to test the significance of their differences. If the new average ability levels across groups were significantly different, then we can claim that the English adaptive system based on the proposed approach delivered the same testing effect as the TOEFL did, but in a semantic representation.

For a small sample such as that in our experiment, an independent t-test should be used to test group differences. However, TOEFL test scores results generally show a non-normal distribution, such as the 2004-2005 test scores in Figure 9 (TOEFL, 2005). Therefore, the non-parametric method of the Mann-Whitney test was used instead in our analysis.

![2004-2005 TOEFL Test Scores Distribution](image)

*Figure 9. 2004-2005 TOEFL scores distribution*

In addition to validate the group discriminations through the independent Mann-Whitney test, the associations between the matched personal TOEFL level and our test level are also calculated. The non-parametric method of Spearman rank-order correlation coefficient, rho (Zikmund, 1997) was used for obtaining the correlations between the average listening, grammar, and reading levels of the two approaches.

To acquire a deeper understanding about the background and experience of the subjects while using the prototype system, during the 20-minutes post-experimental interview, the subjects’ computer-attitude and personal information were also collected. To better utilize the short time of the interview, a semi-structured questionnaire was required to conduct the interview sessions.
Test item pool. Since the benchmark test is against the TOEFL standard, the test item pool needs to contain TOEFL-like objectives and corresponding items. We acquired a 900-item test pool from the English department of I-Shou University in Taiwan, which was used to experiment a fixed-length test of pencil-and-paper format for classifying English abilities of the freshmen. Like the TOEFL test, the item pool contains 300 items each for Grammar, Reading, and Listening. All three objectives were further classified into six difficulty levels: 1 (beginner), 2 (novice), 3 (median), 4 (proficient), 5 (advanced), and 6 (most advanced). These 900 items and 6 difficulty levels are simply taken as they were used at the I-Shou University.

Parameters settings. Frick (1989) indicated that conservative Type I and Type II error rates, $\alpha$ and $\beta$ values, will achieve relative effective judgmental result as explained in Section 2.2. This prototype system adopted 0.025 as suggested in Frick’s research, and thus $\text{LBM} = \frac{1-\beta}{\alpha} = 39$ and $\text{UBN} = \frac{\beta}{1-\alpha} = 0.02564$. As for the probabilities of responding correctly on a test item for the mastery examinee ($P_m$) and for the non-mastery examinee ($P_{nm}$), we took the suggested values of 0.85 and 0.55, respectively, from the domain experts at I-Shou University. The rationale is that a mastery examinee at any ability level is similar to the traditionally perceived score in class, which is around 85 or higher in Taiwan. A current practice in Taiwan for mapping grade A as used in the U.S.A. holds a numerical score of 80 or above. However, it is not difficult for the average students to get an 80 due to different evaluation methods used in classes. Therefore, 85 is commonly recognized by Taiwan’s college teachers, including the authors, as a true mastery-level score that is equivalent to grade A. On the other hand, the perceived non-mastery is less than 60. To avoid being criticized the insignificantly different from just over 60, 55 is chosen in this study. They are similar to 85 and 40 selected by Frick (1989) in his if-then rules.

Stopping condition. Frick (1989) pointed out that for any mastery or non-mastery examinee, there is a high chance of completing the adaptive testing process with a smaller number of items. However, it may consume all items in the item pool before a medium-ability examinee could be decided. To avoid such a situation, we performed a simple simulation using the random number generators in Microsoft Excel 2000 to simulate 40 adaptive testing sessions. Based on the Central-Limit Theorem, for a large sample size such of 30 and above, the sampling distribution will have the same mean as the population, but the variance will be divided by the sample size (Tijms, 2004). Accordingly, the 40 sessions held in this simulation are adequate for obtaining the point estimate. As seen in Table 2, two of the extreme situations of $P=0.55$ and $P=0.85$, as discussed above, are compared. The mastery sessions ($P=0.85$) required less number of items before determination, while the non-mastery sessions ($P=0.55$) needed an average of 16 items to reach a decision. To be conservative, we set the stopping condition to be 25 items for an undecided result.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$P=P_{nm}=0.55$</th>
<th>$P=P_m=0.85$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sessions</td>
<td>Mastery</td>
<td>Non-mastery</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>Average number of items</td>
<td>16</td>
<td>13.9</td>
</tr>
</tbody>
</table>

Note: 40 sessions

<table>
<thead>
<tr>
<th>Group</th>
<th>Score range (#)</th>
<th>Group</th>
<th>Score range (#)</th>
<th>Group</th>
<th>Score range (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GroupR1</td>
<td>1~10 (11)</td>
<td>GroupL1</td>
<td>1~10 (8)</td>
<td>GroupG1</td>
<td>1~10 (15)</td>
</tr>
<tr>
<td>GroupR3</td>
<td>16~20 (13)</td>
<td>GroupL3</td>
<td>16~20 (15)</td>
<td>GroupG3</td>
<td>16~20 (17)</td>
</tr>
<tr>
<td>GroupR5</td>
<td>26~30 (16)</td>
<td>GroupL5</td>
<td>26~30 (9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. TOEFL scores by group segments

Data analysis and discussion

In the TOEFL CBT test, Reading, Listening, and Grammar each have 30 points, where Composition has 6 points that are included in Grammar’s 30 points. Although Taiwan has implemented the TOEFL CAT test for some years already, a few subjects attained the old TOEFL test scores, which were mapped to the new test scores before the analysis. Since Reading and Listening each has 30 points, the subjects were divided into six groups based on their
scores in equivalent distances. Grammar was divided into five groups since it only has 24 points excluding the 6 points from Composition. However, as seen in Figure 9, the lower test scores on all three categories have much smaller percentages than the rest. We merged the lowest two groups, 1-5 and 6-10, into one group to have a better sample count distribution, as seen in Table 3.

In the first step, the independent Mann-Whitney test was used to validate whether or not the group segments could significantly discriminate the average group TOEFL scores. As seen in Tables 4-6, all but one P value was smaller than 0.01, which means that the group segments were divided correctly in terms of the TOEFL scores.

### Table 4. Mann-Whitney test on TOEFL reading scores

<table>
<thead>
<tr>
<th>Comparison code</th>
<th>Pair-wise groups</th>
<th>Number</th>
<th>Mean rank</th>
<th>U-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOEFLr1</td>
<td>GroupR5</td>
<td>12</td>
<td>22.5</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupR4</td>
<td>16</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLr2</td>
<td>GroupR4</td>
<td>16</td>
<td>21.5</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupR3</td>
<td>13</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLr3</td>
<td>GroupR3</td>
<td>13</td>
<td>14.31</td>
<td>9.000</td>
<td>.002**</td>
</tr>
<tr>
<td></td>
<td>GroupR2</td>
<td>8</td>
<td>5.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLr4</td>
<td>GroupR2</td>
<td>8</td>
<td>15.5</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupR1</td>
<td>11</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** Significant at P<0.01; ** Significant at P<0.05; * Significant at P<0.1

### Table 5. Mann-Whitney test on TOEFL listening scores

<table>
<thead>
<tr>
<th>Comparison code</th>
<th>Pair-wise groups</th>
<th>Number</th>
<th>Mean rank</th>
<th>U-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOEFLl1</td>
<td>GroupL5</td>
<td>9</td>
<td>21</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupL4</td>
<td>16</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLl2</td>
<td>GroupL4</td>
<td>16</td>
<td>23.5</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupL3</td>
<td>15</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLl3</td>
<td>GroupL3</td>
<td>15</td>
<td>20</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupL2</td>
<td>12</td>
<td>6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLl4</td>
<td>GroupL2</td>
<td>12</td>
<td>14.5</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupL1</td>
<td>8</td>
<td>4.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** Significant at P<0.01; ** Significant at P<0.05; * Significant at P<0.1

### Table 6. Mann-Whitney test on TOEFL grammar scores

<table>
<thead>
<tr>
<th>Comparison code</th>
<th>Pair-wise groups</th>
<th>Number</th>
<th>Mean rank</th>
<th>U-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOEFLg1</td>
<td>GroupG4</td>
<td>6</td>
<td>20.5</td>
<td>.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupG3</td>
<td>17</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLg2</td>
<td>GroupG3</td>
<td>17</td>
<td>30.47</td>
<td>9.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupG2</td>
<td>22</td>
<td>17.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOEFLg3</td>
<td>GroupG2</td>
<td>22</td>
<td>26.5</td>
<td>0.000</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>GroupG1</td>
<td>15</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** Significant at P<0.01; ** Significant at P<0.05; * Significant at P<0.1

### Table 7. Mann-Whitney test on SPRT-based listening ability levels

<table>
<thead>
<tr>
<th>Comparison code</th>
<th>Pair-wise groups</th>
<th>Number</th>
<th>Mean rank</th>
<th>U-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPRT11</td>
<td>GroupL5</td>
<td>9</td>
<td>17.72</td>
<td>29.5</td>
<td>.014**</td>
</tr>
<tr>
<td></td>
<td>GroupL4</td>
<td>16</td>
<td>10.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRT12</td>
<td>GroupL4</td>
<td>16</td>
<td>19.06</td>
<td>71</td>
<td>.049**</td>
</tr>
<tr>
<td></td>
<td>GroupL3</td>
<td>15</td>
<td>12.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRT13</td>
<td>GroupL3</td>
<td>15</td>
<td>16.8</td>
<td>48</td>
<td>.038**</td>
</tr>
<tr>
<td></td>
<td>GroupL2</td>
<td>12</td>
<td>10.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRT14</td>
<td>GroupL2</td>
<td>12</td>
<td>12.79</td>
<td>20.5</td>
<td>.031**</td>
</tr>
<tr>
<td></td>
<td>GroupL1</td>
<td>8</td>
<td>7.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** Significant at P<0.01; ** Significant at P<0.05; * Significant at P<0.1
In the second step, based on the TOEFL segments, we calculated the average group ability levels collected from our prototype CAT system. Similarly, the Mann-Whitney test was used to examine the significant differences of the average group ability levels. Tables 7-9 showed that except for the SPRTr4 in Reading, most other group comparisons were significantly different at P-values smaller than 0.05. We looked carefully at the raw data and found that there was an outlier in GroupR2. Upon removal of this outlier, the Mann-Whitney test gave a P value of .048 which matched with the overall performance as seen in the parentheses in Table 8.

**Table 8. Mann-Whitney test on SPRT-based reading ability levels**

<table>
<thead>
<tr>
<th>Comparison code</th>
<th>Pair-wise groups</th>
<th>Number</th>
<th>Mean rank</th>
<th>U-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPRTr1</td>
<td>GroupR5</td>
<td>12</td>
<td>17.92</td>
<td>55</td>
<td>.05 *</td>
</tr>
<tr>
<td></td>
<td>GroupR4</td>
<td>16</td>
<td>11.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRTr2</td>
<td>GroupR4</td>
<td>16</td>
<td>17.34</td>
<td>66.5</td>
<td>.094 *</td>
</tr>
<tr>
<td></td>
<td>GroupR3</td>
<td>13</td>
<td>12.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRTr3</td>
<td>GroupR3</td>
<td>13</td>
<td>13.19</td>
<td>23.5</td>
<td>.037 **</td>
</tr>
<tr>
<td></td>
<td>GroupR2</td>
<td>8</td>
<td>7.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GroupR2</td>
<td>8 (7)</td>
<td>12.38 (12.57)</td>
<td>25 (17)</td>
<td>.111</td>
</tr>
<tr>
<td></td>
<td>GroupR1</td>
<td>11</td>
<td>8.27 (7.55)</td>
<td></td>
<td>(.048**)</td>
</tr>
</tbody>
</table>

*** Significant at P<0.01; ** Significant at P<0.05; * Significant at P<0.1

**Table 9. Mann-Whitney test on SPRT-based grammar ability levels**

<table>
<thead>
<tr>
<th>Comparison code</th>
<th>Pair-wise groups</th>
<th>Number</th>
<th>Mean rank</th>
<th>U-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPRTg1</td>
<td>GroupG4</td>
<td>6</td>
<td>16.92</td>
<td>21.5</td>
<td>.033 **</td>
</tr>
<tr>
<td></td>
<td>GroupG3</td>
<td>17</td>
<td>10.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRTg2</td>
<td>GroupG3</td>
<td>17</td>
<td>24.65</td>
<td>108</td>
<td>.022 **</td>
</tr>
<tr>
<td></td>
<td>GroupG2</td>
<td>22</td>
<td>16.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRTg3</td>
<td>GroupG2</td>
<td>22</td>
<td>23.64</td>
<td>63.5</td>
<td>.001 ***</td>
</tr>
<tr>
<td></td>
<td>GroupG1</td>
<td>15</td>
<td>12.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** Significant at P<0.01; ** Significant at P<0.05; * Significant at P<0.1

The Spearman’s rho calculated the average of 60 subjects’ focusing on grammar, reading, and listening levels between the TOEFL scores and our test scores which were declared as .685, .704, and .775, respectively. Note that these three rhos are all significant at 0.01 level. This indicates our test scores were matched well with the TOEFL scores within the grammar, reading, and listening categories.

The subject profile reveals that majority of our subjects are computer literate and had experiences with computer-based tests: 55% of them are male, 80% of them like to use computers, 75% of them assessed their computer skill to be above average, and 72% of them actually took the computer-based TOEFL. Out of the 15 computer-attitude questions covered in the 20 minutes after-test interview, only an average of 3 questions were found negative towards computer-based tests. The reason being, feeling nervous, not concentrating enough, and feeling stressed while taking a computer-based adaptive test. In general, it is still a challenge for a computer-based test, especially the adaptive format that induces a heavier stress to the examinees. On the other hand, the subjects’ feedback on the acceptance of the user interface was positive and the trust in the computer-based test is an increasing potential momentum accepted by the examinees.

**Conclusions and implications**

In order to promote the various benefits of adaptive testing in classroom and business daily KSAO routines, we proposed an SPRT-based CAT approach that would not only avoid the tedious pre-test procedure for maintaining the item pool but also produce a more distinguishable level of the examinee’s ability instead of a pass-or-fail decision. Furthermore, we have demonstrated its technical feasibility by implementing an English CAT prototype system as well as its criterion validity by a benchmark test against the well-validated TOEFL test.
One observation from the experiment is that the P-value level of the SPRT-based score differences are around 0.05, while those for TOEFL are 0.01, which is also supported by their correlations that range from 0.685 to 0.775, a significant but not perfectly-matched level. According to our after-test interviews, many subjects admitted that they were nervous since they did not prepare for this online English test as they did for TOEFL which, in many cases, was taken by the examinees years ago. Nevertheless, this situation may actually reflect their true English abilities. Moreover, the N-tier SPRT item pool was not as rigorously tested against IRT 3-PLM as suggested by Kalohn & Spray (1999) for certified or licensed centers, which may be another factor. Despite the above obstacles, the benchmark test with only 60 subjects is in general very positive towards this criterion-validity test against TOEFL. Consequently, we are confident that the proposed approach can effectively evaluate the student or employee KSAO’s appropriately while meeting various assessments needs flexibly under a small-scale scenario.

There are two major implications. First, because this SPRT-based approach is sound in theory, effective in performance, and efficient in implementation, it can become an IT-proven innovation strategy as a cost-effective tool. In other words, this approach may prolong the life cycle of an adaptive testing system by having the HR staff, the domain experts/teachers, and the employee or job applicants/students focus only on their designated routines in process administration, item pool maintenance, and test taking, respectively, without additional IT or adaptive-testing expertise.

Another implication is that certified, license centers, or even software developers can provide this adaptive testing approach as a new business model to their business clients and school teachers. That is, an application service provider (ASP) leases the testing software and/or system as a package to school teachers or business clients, and lets them operate and maintain the testing process. In this scenario, adaptive testing systems can be more affordable and flexible in meeting academia and business versatile needs, which provide a good motivation and win-win situations for both the ASP and the adopters.

Despite our effort in trying to carefully conduct this study, there is one significant limitation to be noted while adopting the conclusions and implications. That is, although the sample size of 60 professionals is adequate to claim the feasibility of the proposed SPRT-based CAT approach, a more robust experimental result would require a larger sample size compared to 60. One immediate future work is to conduct a field test of the proposed N-tier adaptive testing model, which is made possible by the Teaching Excellence project grant awarded by the Ministry of Education in Taiwan, R.O.C. to the first author’s University. It is in the process of implementing this adaptive testing system in the school of Management, which provides a natural testbed for the future field study.

Acknowledgements

This research project was partially sponsored by Ministry of Education of the Republic of China with grant number H-041. The authors would like to thank the anonymous referees for their valuable suggestions in improving the readability of this paper.

References


ETS (2001). A study of the use of collateral statistical information in attempting to reduce TOEFL® IRT item parameter estimation sample sizes, Report RR-01-11, TOEFL-TR-17, retrieved June 2, 2008, from


Comparison of Dimension Reduction Methods for Automated Essay Grading

Tuomo Kakkonen, Niko Myller, Erkki Sutinen and Jari Timonen
Department of Computer Science and Statistics, University of Joensuu, Finland // niko.myller@cs.joensuu.fi // Tel. +358 13 251 7929 // Fax. +358 13 251 7955

ABSTRACT

Automatic Essay Assessor (AEA) is a system that utilizes information retrieval techniques such as Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), and Latent Dirichlet Allocation (LDA) for automatic essay grading. The system uses learning materials and relatively few teacher-graded essays for calibrating the scoring mechanism before grading. We performed a series of experiments using LSA, PLSA, and LDA for document comparisons in AEA. In addition to comparing the methods on a theoretical level, we compared the applicability of LSA, PLSA, and LDA to essay grading with empirical data. The results show that the use of learning materials as training data for the grading model outperforms the k-NN-based grading methods. In addition to this, we found that using LSA yielded slightly more accurate grading than PLSA and LDA. We also found that the division of the learning materials in the training data is crucial. It is better to divide learning materials into sentences than paragraphs.

Keywords
Automatic essay grading, Dimensionality reduction, Latent semantic analysis, Probabilistic latent semantic analysis, Latent Dirichlet allocation

Introduction

In this paper, we compare in the context of automated essay grading three well-known dimensionality reduction methods, namely Latent Semantic Analysis (LSA) (Deerwester et al., 1990; Landauer et al., 1998), and related statistical models Probabilistic LSA (PLSA) (Hofmann, 2001) and Latent Dirichlet Allocation (LDA) (Blei et al., 2003), all of which are commonly used in information retrieval (IR). All of these techniques assume that one can model documents as mere collections of words without giving any consideration to word order. Because these methods try to disambiguate words with multiple meanings (i.e. polysemes) and to find groups of words that belong to the same context or that have the same meaning (i.e. synonyms), they are suitable for comparing similarities between two documents at the semantic level rather than simply at the level of keywords. These three methods all take a data-driven approach to the problem of learning the meanings of words. Because the meanings are based on the contexts of word usages, no thesauri or dictionaries are needed in the process.

It is possible to automatically grade essays (i.e. free-text responses in examinations) by comparing them to select learning materials and human-graded essays. We used these learning materials as a comparative basis for determining the amount of relevant content in an essay. In order to fine-tune this method so that it produces grades that correspond to those given by human assessors, it is necessary first to train the system with human-graded essays. Dimensionality reduction refers to the process of giving individual words in the model weights that correspond to their significance in the context of the topic as a whole. The purpose of the dimensionality reduction step is to reduce the noise and unimportant details in the data so that the underlying semantic structure can be used to compare the content of essays.

Since the dimensionality reduction problem in automated essay grading is intuitively clear, we can state it in the following way. The more we make use of pure word occurrences, the more we emphasize various details (such as completely irrelevant words in the corpus) as the basis for grading essays. But what is really required is to focus on the most important of those words that embody the conceptual content of the learning corpus. This sort of procedure is in fact similar to the kind of work that is performed by a human assessor. As she marks, a human assessor identifies important concepts but skims over or pays less attention to the peripheral content of whatever essay she may be marking. One could characterize this process of paying less attention to peripheral contents in manual essay grading in IR-terms as dimensionality reduction.

Even though LSA performs well in various IR tasks apart from automatic essay grading, it is characterized by some undesirable characteristics that we will discuss later in this article. In order to resolve the problems that appear in
LSA, new probabilistic models such as PLSA and LDA have been proposed. While PLSA and LDA have been shown to produce better results in IR tasks (Blei et al., 2003; Hofmann, 2001; Yu-Seop et al., 2002), their performance has not yet been tested in automated essay grading. We have therefore implemented them as a part of our essay grading system.

We will begin with a brief review of earlier work on automatic essay grading and on LSA, PLSA and LDA. This is followed by an introduction to the architecture of the essay grading system, Automatic Essay Assessor (AEA). Descriptions of LSA, PLSA and LDA in essay grading domain can be found in subsequent sections. We will report on the results of an empirical comparison of the methods in following section. Finally, we will offer our conclusions and indicate the research that we still intend to carry out in the field of automated essay grading and IR.

**Previous Work**

**Text Categorization and Dimensionality Reduction**

Automatic essay grading is closely related to automatic text categorization, which has been researched since the 1960s. Many of the methods that are used in other IR tasks have been found to be applicable to text categorization. Thus, for example, support vector machines (SVMs) and dimensionality reduction methods such as LSA, have been successfully applied to the problem (see Sebastiani (2002) for a detailed overview of research into text categorization).

LSA, PLSA and LDA have all been demonstrated to be reliable methods in IR. The earliest model LSA (also known as Latent Semantic Indexing (LSI)) has been successfully applied to various IR tasks from information filtering (Foltz and Dumais, 1992) and classification (Zelikovitz and Hirsh, 2001) to image retrieval (Praks et al., 2003). Early applications of PLSA and LDA also include document retrieval and classification (Blei et al., 2003; Hofmann, 2001; Brants, 2005).

Several studies have shown that PLSA outperforms LSA in document modeling and classification tasks (Hofmann, 2001; Yu-Seop et al., 2002). When Hofmann (2001) compared the performance of the two methods in document modeling by measuring the perplexity of the language models and by comparing them to a unigram baseline model, he found that the improvement of the standard LSA model was less than a factor of two whereas PLSA yielded improvements of a factor of more than three.

Similar results were observed on a document retrieval task performed with the four standard medium-size document collections, MED, CRAN, CACM, and CISI. In these cases, the performance of LSA and PLSA was compared against a baseline term frequencies-based vector space model. While LSA improved the precision results by 0.8% and 16.7% on two of the test sets, it delivered lower precision rates with the two other test sets. PLSA improved the precision results 14.7%...58.3% on all the test sets. The difficulties encountered during the selection of the optimal dimension for LSA is a well-reported problem (Landauer et al., 1998; Bingham and Mannila, 2001; Globerson and Tishby, 2003), and the selection is typically performed by ad hoc heuristics. Blei et al. (2003) have shown that LDA outperforms PLSA in document classification and collaborative filtering tasks with medium-size collections.

**Automatic essay grading**

Research to develop computer systems for automatic essay grading has been carried out since the 1960s and several approaches have been proposed. Project Essay Grade (PEG) (Page, 1966; Page and Petersen, 1995) uses multiple regression techniques, and Larkey applies text categorization techniques and linear regression methods. Bayesian essay testing system (BETSY) (Rudner and Liang, 2002) is based on Bayesian Networks, Intelligent Essay Assessor (Foltz et al., 1999) uses LSA for the content analysis of essays, and E-RATER (Burstein et al., 1998) is a hybrid system that combines Natural Language Processing (NLP) and statistical techniques.

An automatic grading module is central to any essay assessment system. Content-based grading can be performed by means of two methods: (1) by comparing an essay to human-graded essays and assigning the grade based on the
grades of the \(k\) nearest neighboring essays, and (2) by basing the grading on both human-graded essays and course content. Of the three methods discussed in this article, only LSA has been previously used for automatic grading and other educational applications such as an intelligent tutoring system (Wiemer-Hastings et al., 1999) and for assessing student summaries (Wade-Stein and Kintsch, 2003). LSA has proved to be one of the most successful methods for content-based essay grading. Depending on the test set, Landauer et al. (1997) and Foltz et al. (2000) have, for example, reported correlations from 0.64 to 0.84 between grades given by two human assessors and correlations from 0.59 to 0.89 between the LSA-based grading system and human graders. This means that LSA-based systems perform as well as the human graders. Our aim was to determine whether the use of PLSA and LDA in our grading system would improve the accuracy of the grading and help us to avoid the problems that are characteristic of the LSA model.

According to Kaplan, Wolff, Burstein, Li, Rock, and Kaplan, the four quality criteria for an automated essay grading system are *accuracy, defensibility, coachability* and *cost-efficiency*. For a system to be acceptable, it must deliver on all these criteria. An accurate system is capable of producing reliable grades measured by the correlation between a human grader and the system. In order to be defensible, the grading procedure employed by the system must be traceable and educationally valid. In other words, it should be possible rationally to justify and explain the grading method and the criteria for given grades. Coachability refers to the transparency of the grading method. If the system is based on simple, surface-based methods that ignore content, students could theoretically train themselves to circumvent the system and so obtain higher grades than they deserve. It is also self-evident that an automated grading system must be cost-efficient because its ultimate purpose is to reduce the total costs of assessment.

Of the four requirements, accuracy is most readily accounted for by all the systems represented in Table 1. Results reported as early as the nineteen sixties indicated that automated grading systems can grade essays as accurately as humans (Page, 1966). Since the cost-effectiveness of the systems is dependent on the number of graded essays that are graded, it would not make sense to collect a set of several hundred human-graded essays in order to grade just a few. It is from this point of view that the methods based on LSA are most effective because LSA can apply course materials as a basis for assessment.

The most problematic requirement for the current systems is probably coachability. If a writer is familiar with the grading principles of the system, he can mislead the system to render better grades than are deserved. One possible solution to this problem is to automatically identify ‘suspicious’ essays and give them to a human assessor for checking.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Defensibility</th>
<th>Coachability</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PEG</strong></td>
<td>Grades as accurate as human graders measured by the correlation.</td>
<td>Relies heavily on measuring surface features. But the scoring principles are easily traceable.</td>
<td>Because the applied measures are simple, coachability can cause problems.</td>
<td>Relatively large number of pre-graded essays are needed, but it is less costly to compute than, for example, LSA.</td>
</tr>
<tr>
<td><strong>TCT</strong></td>
<td>Like PEG.</td>
<td>Uses both surface features and content measures, and is thus more defensible than PEG.</td>
<td>Does not pay attention to the structure.</td>
<td>Like PEG.</td>
</tr>
<tr>
<td><strong>BETSY</strong></td>
<td>Like PEG.</td>
<td>Measures only the content.</td>
<td>Measures only the content.</td>
<td>Like PEG.</td>
</tr>
<tr>
<td><strong>LSA</strong></td>
<td>Like PEG.</td>
<td>Like BETSY.</td>
<td>Like BETSY.</td>
<td>In addition to pre-scored essays, course material can be used for training.</td>
</tr>
<tr>
<td><strong>E-RATER</strong></td>
<td>Like PEG.</td>
<td>One of the design principles. The strongest of the systems in this respect.</td>
<td>Also analyzes the structure and organization.</td>
<td>270 essays are required for training.</td>
</tr>
</tbody>
</table>

*Table 1. Comparison of five essay assessment methods according to the criteria proposed by Kaplan et al. (1998)*
Automatic Essay Assessor: AEA

Our particular grading system, *Automatic Essay Assessor* (AEA), is designed to grade essays automatically (Kakkonen and Sutinen, 2004). It bases grading on a scoring model that has been created by comparing a set of manually graded essays to course materials. While this system is not limited to a particular language, it currently supports only Finnish. It is also possible, in addition, to apply several dimensionality reduction methods for modeling and measuring the content similarities between documents. The system currently contains implementations of LSA, PLSA (Kakkonen et al., 2005a), and LDA (Kakkonen et al., 2006), and this is discussed in more detail in the following section below. The structure of AEA is represented in Figure 1.

![Figure 1. The architecture of AEA](image)

Figure 1 shows how the system logically consists of two main parts: analysis and grading phases. In the analysis phase, the *scoring model* is created from the course materials (passages from course textbooks, lecture notes and pre-graded essays). The text passages are selected from the course materials on the basis of their relevance to the essay prompt; all the passages that the students are supposed to have read in order to be able to answer the question successfully are used as the bases for the evaluation. In the grading phase, the essays are graded according to the scoring model created in the analysis phase.

There are three main components at the implementation level. These are NLP, the grade definition, and the dimension reduction components. The *NLP component* consists of a syntactic parser and preprocessor. We use the Constraint Grammar Parser for Finnish (Lingsoft Inc., 2007) in order to produce the base forms of words in the input documents. The *preprocessor* performs the standard preprocessing stages that are used in LSA, PLSA and LDA, namely, stopword removal and entropy-based term weighting. If AEA is applied to languages other than Finnish, the parser and stopword lists must be replaced with relevant ones. We used the NLP component for two distinct tasks: firstly, to create a *word-by-context matrix* (WCM) representing the course materials, and, secondly, to build document vectors of the *pre-scored essays* and essays to be graded. *Pre-scored essays* in this context refer to those essays that have been given a grade by the teacher and that are used for calibrating the AEA before the actual
grading. Because WCM contains the number of occurrences of each word in each context in the course materials (i.e. document contexts, the paragraph contexts, or the sentence contexts), it is indeed a collection of document vectors from all possible contexts.

Scoring model creation progresses in the following way. Firstly, the WCM (representing the course materials) is processed in the dimensionality reduction component with LSA, PLSA, or LDA in order to create a reduced-dimensional representation of the WCM. The reason for this step is to reduce the noise in the WCM. This allows the unimportant details to dissipate and the underlying semantic structure to become more patent. (While issues relating to these dimensionality reduction methods will be further discussed in the next section, a more detailed description of the scoring model creation can be found in Kakkonen et al. (2005a) and Kakkonen et al. (2005b).)

The next step is to compare the document vectors that have been created from the manually graded essays to the reduced WCM so as to determine the similarities between each essay and the course materials. This comparison between an individual student essay and the reduced WCM refers to the process whereby the distances between a document vector of an essay and every document vector in the reduced WCM are calculated and summed to form the similarity value between the essay and the WCM. In order to do this, we apply the cosine between the document vectors as a measure of their semantic similarity. On the basis of these similarities, the predefined grade categories, \( g_0, g_1, \ldots, g_{G-1} \), are associated with similarity value limits, \( l_0, l_1, \ldots, l_{G} \), where \( G \) is the number of grade categories, and \( l_G = \infty \) and normally \( l_0 = 0 \) or \( l_{G} = -\infty \) and \( l_j \) (\( 0 < j < G \)) are, for example, the average of similarity values of training essays in the grade categories \( g_{j-1} \) and \( g_j \).

The next step is the grading phase. In this phase, the NLP component creates document vectors from each of the essays to be graded. Grades are determined with the scoring model consisting of the reduced WCM and limits for the grade categories. The essay’s document vector \( d \) is compared to WCM in order to define the similarity value as described above. An essay is assigned to a single grade category on the basis of the similarity value, \( \text{sim}(d) \), and the limits of grade categories as follows: \( l_i < \text{sim}(d) \leq l_{i+1} \rightarrow d \in g_i \).

The scoring model is then validated by measuring the Spearman correlation between the grades given to test essays by the system those given by the human assessors. This phase is essential (especially when applying LSA) because the selection of dimensionality in the method is somewhat arbitrary.

**Dimensionality Reduction Methods in AEA**

Dimensionality reduction in AEA refers to the process by means of which the individual words that will be used for comparing essays with learning materials are assigned a weight according to their significance. In order to determine the optimal parameters for dimensionality reduction, one needs to train AEA with different parameters and to predict the corresponding grades for essays by using each of these models (Kakkonen et al., 2005b). We shall now review the three dimensionality reduction methods (LSA, PLSA and LDA) that are used by AEA. The results of this comparison of LSA, LDA and PLSA are reported in later section.

**Latent Semantic Analysis**

Dimensionality reduction in LSA is based on singular value decomposition (SVD), a form of factor analysis. In LSA the original WCM is approximated by decreasing the number of singular values in the SVD of WCM. This has been shown to increase the dependencies between contexts and words (Deerwester et al., 1990; Landauer et al., 1998).

While LSA offers a feasible method to compare the similarities between two documents, there are some problems inherent in this method. The number of singular values in the dimensionality reduction is usually selected by means of an ad hoc heuristics. While dimensionality reduction can result in a reduced WCM that contains negative values, this is not necessarily a problem because it allows the document vectors to be defined in a larger subspace (by comparison with a subspace that allows only vectors with positive components), and this could be helpful, especially when LSA is used with small training sets. In spite of this, the definition of the WCM becomes problematic because
it is unclear what the meaning of a context with a negative number of words is. This would, in addition, prohibit the use of the reduced WCM to define probability distributions. Problems with dimensionality reduction have also been reported in other LSA-based applications (Landauer et al., 1998; Bingham and Mannila, 2001; Globerson and Tishby, 2003).

The selection of dimensionality in LSA is fraught with problems (Bingham and Mannila, 2001; Globerson and Tishby, 2003; Landauer et al., 1998). Kim et al. (2003) concluded on the basis of their experiments that LSA and PLSA dimensionality do not have a specific linkage to semantic knowledge construction. Kim et al. (2003) were unable, in other words, to find the constant dimension that fits their experimental data. This implies that a single dimension is not applicable to all essay collections in AEA. The problems generated by dimensionality reduction led us to develop ADRM, an automatic dimensionality reduction method that searches case-specifically for the dimension that best fits each set of essays (Kakkonen et al., 2005b). But since we have not yet modified ADRM for PLSA and LDA, the experiments reported in these articles were run on the standard LSA that repeats the scoring for all the possible dimensions.

Probabilistic Latent Semantic Analysis

Although we found that the practical performance of LSA is good in many IR tasks as well as in essay grading, it has been discovered that it possesses additional flaws (Quesada, 2003; Hofmann, 2001) apart from the dimensionality selection problem. LSA, for example, does not define a proper probability distribution, and, even more seriously, the reduced matrix can contain negative values. To solve these problems, Hofmann (2001) proposed PLSA, a probabilistic extension to LSA, which we had already utilized PLSA in automated essay grading (Kakkonen et al., 2005a). PLSA is based on a statistical model that is referred as an aspect model. An aspect model is a latent variable model for co-occurrence data, which associates unobserved class variables \( z_k \), \( k \in \{1,2,...,K\} \), with each observation, where \( K \) is the number of latent classes. While the number of latent classes is an important parameter that needs to be selected in the same way as LSA, there are, as we will point out below, several existing solutions to this problem. In our settings, the observation is an occurrence of a word \( w_j \), \( j \in \{1,2,...,M\} \), in a particular document/context \( d_i \), \( i \in \{1,2,...,N\} \), as in WCM, where \( M \) is the number of words and \( N \) is the number of documents in the collection. Latent classes can be understood as the topics that comprise the text. The probability distributions that associate the latent variables with words and documents describe how closely they are associated with each topic. The generative model for the observation is defined as follows:

1. Obtain a document \( d_i \) in which a word occurrence will be observed with probability \( P(d_i) \).

2. When the document \( d_i \) is known, select the topic \( z_k \) of the word with probability \( P(z_k \mid d_i) \). This probability distribution is also a measure of the extent to which the document is relevant to each topic.

3. When the topic is known, select a word \( w_j \) whose occurrence is observed with probability \( P(w_j \mid z_k) \).

In this way the observation pair \( (d_i, w_j) \) can be formulated so that the latent class variable can be summed out. Equation (1) shows the probability of observing a pair \( (d_i, w_j) \).

\[
P(d_i, w_j) = P(d_i)P(w_j \mid d_i), \text{ where } P(w_j \mid d_i) = \sum_{k=1}^{K} P(w_j \mid z_k)P(z_k \mid d_i) \quad (1)
\]

When one uses PLSA in essay grading or IR, one’s first task is to construct the model. This means approximating the probability mass functions from the training data with machine learning techniques. In our own case we used the comparison materials consisting of the assignment of specific texts as training material. Hofmann (2001) proposes the use of the Expectation Maximization algorithm to identify the parameters for the probability mass function from the training materials.
The *Expectation Maximization (EM)* algorithm can be used to build a model with a maximum likelihood formulation of the learning task (Dempster et al., 1977). In EM, the algorithm alternates between the following two steps: (i) an *expectation (E)* step in which posterior probabilities are computed for the latent variables on the basis of the current estimates of the parameters (see Equation (2)), and (ii) a *maximization (M)* step in which parameters are updated on the basis of the minimization criteria and in dependence on the posterior probabilities computed in the E-step (see Equations (3) and (4)).

\[
P(z_k | d_i, w_j) = \frac{P(w_j | z_k)P(z_k | d_i)}{\sum_{j=1}^{K} P(w_j | z_i)P(z_i | d_i)} \tag{2}
\]

\[
P(w_j | z_k) = \frac{\sum_{n=1}^{N} n(d_i, w_j)P(z_k | d_i, w_j)}{\sum_{m=1}^{M} \sum_{n=1}^{N} n(d_i, w_m)P(z_k | d_i, w_m)} \tag{3}
\]

\[
P(z_k | d_i) = \frac{\sum_{j=1}^{N} n(d_i, w_j)P(z_k | d_i, w_j)}{\sum_{m=1}^{M} n(d_i, w_m)} \tag{4}
\]

In the equations, \( n(d_i, w_j) \) stands for the count of the word \( w_j \) in the document \( d_i \).

The standard EM algorithm can, however, overfit the model to the training data and thus perform poorly with unseen data. Because this algorithm is iterative and converges relatively slowly, it can increase runtime dramatically, especially with large data sets. Because of this, Hofmann (2001) proposed another approach that he calls *Tempered EM* (TEM), which is a derivation of standard EM algorithm. (We refer the interested reader to Hofmann (2001) for further information.)

In PLSA, a query \( q \), or, in our case, an essay, can be ‘added’ or *folded into* the model by using TEM. When folding in a new essay, the only difference is that the probabilities \( P(w_j | z_k) \) are retained as they are, and only the probabilities \( P(z_k | q) \) are updated during the M-step.

The similarity between a document or comparison material \( d_i \) in the model and an essay \( q \) folded into the model can be calculated in a similar way to LSA, with the cosine of the angle between the vectors containing the probability distributions \( (P(z_k | q))_{k=1}^{K} \) and \( (P(z_k | d_i))_{k=1}^{K} \) (Hofmann, 2001). Another similarity measure is the logarithm of the a posteriori probability for the comparison material passage \( d_i \) given the essay \( q \) formulated by Girolami and Kabán (2003). This is shown in Equation (5).

\[
sim(d_i, q) = \sum_{w_j \in d_i} n(d_i, w_j) \log \sum_{k=1}^{K} P(w_j | z_k)P(z_k | q) \tag{5}
\]

Unlike LSA, PLSA defines proper probability distributions to the documents. PLSA is interpretable with its generative model, latent classes or topics and graphical illustrations in \( K \)-dimensional space (Hofmann, 2001). In IR, PLSA yielded equal or better results when compared to LSA (Yu-Seop et al., 2002; Hofmann, 2001). Hofmann (2001) demonstrated that the accuracy of PLSA can be increased by increasing the numbers of latent variables. The combination of several similarity scores (such as the cosines of angles between two documents) from models with different number of latent variables also increases accuracy. The selection of the dimension is not therefore as crucial as it is in LSA.

A problem with PLSA is that the algorithm used to compute the model – EM or its variant – is probabilistic and it can converge to a local maximum. Hofmann (2001) believes this is not a severe problem and that the differences
between separate runs are small. Baldi et al. (2003) and Si and Jin (2005) are among those who have pointed out that PLSA often overfits the model to the training data. The recent research of Brants (2005) shows that this is not necessarily the case. The generative model of PLSA and the assumptions behind it have also been characterized as flawed by Blei et al. (2003), because there is a need to estimate probability to acquire an unseen document. Blei et al. (2003) solved this problem by proposing another statistical framework, LDA, which we will examine in the next section.

Latent Dirichlet Allocation

The principles of LDA are similar to those of PLSA. The difference is that there is no need in LDA to estimate the probability of obtaining a document, and there is thus no need to perform the difficult estimation process when adding unseen documents to the model. Instead this is achieved by changing the generative model in such a way that it separates the process for each document and uses the word-latent class distribution to determine the document-latent class distribution. LDA assumes the following generative process for each document $d_i$ in a corpus consisting of $N$ documents that contains $M$ distinct words and $K$ distinct latent variables or topics:

1. Choose the length of the document $L$-Poisson($\alpha$). Note that we are (for most of the time) dealing with each sequence of words in a single document and not the distinct words in the corpus. We therefore use indexing $w_i$ for words in a single document $d_i = \{w_1,\ldots,w_L\}$ and $w_m$ for distinct words in a corpus.

2. Choose a parameter vector for the topic distribution $\theta$-Dirichlet($\alpha$), the parameter $\alpha$ is a $K$-vector with components $\alpha_k > 0$ and $\theta$ is a $K$-vector so that $\theta_k \geq 0$ and $\sum_{k=1}^{K} \theta_k = 1$ and $P(\theta|\alpha)$ is the probability density function of the Dirichlet distribution.

3. For each of the $L$ words $w_i$:

   a) Choose a topic $z_i \sim \text{Multinomial}(q)$. Note that $z_i$ or $z_L$ is used when we discuss the topic of each word in a document, and that all topics in a document are referred as $z_{d_i} = \{z_1,\ldots,z_L\}$.

   b) Choose a word $w_i$ from $P(w_i | z_i, \beta)$, a multinomial probability conditioned on the topic $z_i$ where $\beta$ is a $K \times M$ matrix so that $\beta_{ij} = P(w_m | z_k)$ for all $1 \leq m \leq M$ and $1 \leq k \leq K$, where $M$ is the number of distinct words in the corpus.

In order to build up the model for LDA, one should compute the posterior distribution of the latent variables for a given document in the way shown in Equation (6).

$$P(\theta, z_{d_i} | d_i, \alpha, \beta) = \frac{P(\theta, z_{d_i}, d_i | \alpha, \beta)}{P(d_i | \alpha, \beta)} \quad (6)$$

But because Equation (6) is intractable, it needs to be approximated. Blei et al. (2003) introduce an EM-based variational algorithm to approximate the equation and maximize the log likelihood of the model based on the $\alpha$ and $\beta$ parameters. We will describe the algorithm briefly at this point. Further details can be found in Blei et al. (2003) and Minka and Lafferty (2002).

In the E-step, the density function in Equation (6) needs to be approximated with a tractable model. The idea is to minimize the Kullback-Leibner Divergence between the tractable and intractable model by finding the minimal values for Dirichlet parameter $\gamma$ and multinomial parameters $(\phi_1,\ldots,\phi_L)$ in the tractable model. In order to find the
optimal \( \gamma \) and \( \phi \) for each document \( d_i \), Blei et al. (2003) obtained the updated Equations (7) and (8) for these parameters,

\[
\phi_{mk}(d_i) \propto \beta_{km} \exp\{E_p[\log(\theta_k) \mid \gamma(d_i)]\},
\]

where \( E_p[\log(\theta_k) \mid \gamma(d_i)] = \Psi(\gamma(d_i)) - \Psi\left(\sum_{j=1}^{K} \gamma_j(d_i)\right) \) (7)

\[
\gamma_k(d_i) = \alpha_k + \sum_{j=1}^{N} \phi_{nk}(d_i)
\] (8)

Because the \( \gamma \) parameter vector describes the topic distribution for each document, it can be used in a similar way to \( P(z_k \mid d_i) \) in the PLSA model. These two equations are computed repeatedly for all \( l, k \) and \( d_i \) until the lower bound achieved from Jensen’s inequality converges.

In the M-step, \( \alpha \) and \( \beta \) parameters need to be estimated once the new values of \( \phi \) and \( \gamma \) have been calculated. Blei et al. (2003) propose the use of the Newton-Raphson optimization technique to find the stationary point of the \( \alpha \) function by iterating Equation (9). The conditional multinomial parameters \( \alpha \) and \( \beta \) are also updated as in Equations (9) and (10).

\[
a_{new} = a_{old} - H(a_{old})^{-1}g(a_{old})
\] (9)

\[
\beta_{km} \propto \sum_{j=1}^{N} \sum_{l=1}^{L_k} \phi_{lk}(d_i) eq(w_l, w_m),
\] (10)

where \( H(\alpha) \) and \( g(\alpha) \) are the Hessian matrix and gradient respectively at point \( \alpha \) and \( eq(w_l, w_m) \) is 1 if a word \( w_l \) from the document \( d_i \) is the same word as \( m \)th distinct word \( w_m \) in the corpus otherwise 0. After each cycle in the EM algorithm the convergence of the model building is measured by means of the log-likelihood of the model.

A new document or query can be added to the model by using the same procedure. Blei et al. (2003) propose methods for smoothing the distributions in order to avoid zero probabilities when adding new documents that contain previously unseen words.

The \( \gamma \) vector of a document contains the information about how the document belongs to the different latent classes or topics. Because \( \phi \) contains the same information for each word in the document, the similarity between documents can be compared with the same methods that were used in PLSA – by applying either the cosine of the angle between the documents’ \( \gamma \) vectors or the logarithm of the a posteriori probability for the document. Another distance measure that can be used with LDA is the entropic cosine similarity that was formulated by Girolami and Kabán (2003). This is shown in Equation (11).

\[
sim(d_i, q) = \sum_{w_j \in q} n(q, w_j) \log \sum_{k=1}^{K} P(w_j \mid z_k) P(z_k \mid d_i)
\] (11)

PLSA and LDA appear to be very similar methods. Girolami and Kabán (2003) have in fact shown that PLSA may be regarded as a maximum a posteriori maximum likelihood estimate of the theoretical model in LDA – although it is one that uses different assumptions about the distributions. The variational method used in LDA seems to produce lower perplexity language models than PLSA and to perform better in the text categorization tasks (Blei et al., 2003). In spite of this, the variational method has been claimed to overfit the model into the training data in some cases, and another method, namely Expectation Propagation, has been proposed to compute the LDA model in order to achieve lower perplexity (Minka and Lafferty, 2002). But it is not clear how the perplexity correlates with the results in IR or in essay grading. In the next section, we will compare the performances of LSA, PLSA and LDA in essay-grading contexts in order to analyze their performances on this task.
Empirical Comparison of the Methods

The purpose of using the language modeling methods Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA) in the context of automated essay grading is to reduce noise and to compare the similarities of documents (in the case of AEA, the similarities will be between the essays and the course materials). In this research we compared the performance of these methods in order to analyze their differences in these settings. To validate the suitability of our essay grading process (illustrated in Figure 1), we compared it to the *k-Nearest Neighbors* (k-NN) method that is used in essay grading systems (cf. (Larkey, 1998)).

We compared the essay grading accuracy of our implementations of LSA, PLSA and LDA described in section “Automatic Essay Assessor: AEA” with the essay test sets described in Table 2. In Table 2, the column headed *Field* shows the topic of the course (Education, Communication or Software Engineering). The column headed *Train essays* shows the number of training essays that we used for creating the scoring model before grading, and the column headed *Test essays* gives the number of essays that we used for testing the accuracy of the scoring model. We divided the learning materials from the courses either into paragraphs or sentences as shown in the column headed *Div. type*. The total count of text passages (the number of columns in the word-by-context-matrix (WCM)) is recorded in the column headed *No. pass.*, and the column headed *No. words* indicates the total length of the course materials.

The essay sets in Table 2 were graded by a professor and a teacher or a course assistant. The Software Engineering test sets (5 and 6) were graded by two human graders. The correlation between these graders was .88 on all the essays and the correlation between the two human graders for the training essays and the test essays were .89 and .87, respectively.

We ran the experiment with LSA for all the possible dimensions (i.e. from two to the number of contexts in the WCM). In contrast to the number of dimensions in LSA, which is limited to the number of contexts in the WCM, there is no upper limit for the number of latent variables or topics in the PLSA and LDA models. In order to make a fair comparison, we set the same upper limit for the number of variables and topics in PLSA and LDA that we had in LSA. In addition to this we also conducted experiments with a PLSA model (referred to henceforth as PLSA-C) in which the similarity score was defined as the linear combination of similarity values obtained from PLSA models with predefined numbers of latent variables \( K \in \{16,32,48,64,80,96,112,128\} \). When we built up the PLSA models with TEM, we used twenty essays from the training set to test the stopping condition. We utilized the cosine of the angle between the vectors and the logarithm of the a posteriori probability and the entropic cosine as the similarity measures in all the methods where applicable, and selected the highest score.

<table>
<thead>
<tr>
<th>Set No.</th>
<th>Field</th>
<th>Level</th>
<th>Train. essays</th>
<th>Test essays</th>
<th>Grad. scale</th>
<th>Grader</th>
<th>Div. type</th>
<th>No. pass.</th>
<th>No. words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Educ.</td>
<td>Under grad.</td>
<td>70</td>
<td>73</td>
<td>0–6</td>
<td>Prof.</td>
<td>Par.</td>
<td>26</td>
<td>2397</td>
</tr>
<tr>
<td>2</td>
<td>Educ.</td>
<td>Under grad.</td>
<td>70</td>
<td>73</td>
<td>0–6</td>
<td>Prof.</td>
<td>Sent.</td>
<td>147</td>
<td>2397</td>
</tr>
<tr>
<td>3</td>
<td>Comm.</td>
<td>Vocational</td>
<td>42</td>
<td>45</td>
<td>0–4</td>
<td>Course teacher</td>
<td>Par.</td>
<td>45</td>
<td>1583</td>
</tr>
<tr>
<td>4</td>
<td>Comm.</td>
<td>Vocational</td>
<td>42</td>
<td>45</td>
<td>0–4</td>
<td>Course teacher</td>
<td>Sent.</td>
<td>139</td>
<td>1583</td>
</tr>
<tr>
<td>5</td>
<td>Soft. Eng.</td>
<td>Grad.</td>
<td>26</td>
<td>27</td>
<td>0–10</td>
<td>Assist.</td>
<td>Par.</td>
<td>27</td>
<td>965</td>
</tr>
</tbody>
</table>

We applied similar procedures for the k-NN-based grading methods *KNN-LSA*, *KNN-PLSA*, *KNN-PLSA-C*, and *KNN-LDA*. The models were computed by using the training essays alone. We conducted these experiments with the values of \( k \) between 1 and 10 and determined the grade for an essay as a similarity-score-weighted average of the grades of the \( k \) neighboring essays. We reported the correlation for each k-NN method with the value of \( k \) that resulted in the most accurate grading.
Table 3 shows the results of the experiment as measured by means of the Spearman correlation between the grades given by a human assessor and the system. The achieved correlations are comparable to the ones found in the literature. While the correlations for essay sets 5 and 6 vary from .64 to .95, most of the correlations vary between .81 and .95. A comparison with the correlation between the two human graders for the essay sets 5 and 6 (.87) indicates that AEA is able to grade essays as accurately as the human graders (the correlation between the grades by AEA and the course teacher was .90).

The results clearly show that k-NN-based methods are outperformed by the grading method of AEA that uses course materials in the grading process. All the methods yielded the lowest grading accuracy for the Communications test set. We suspect that this was caused by the open-endedness of the essay prompt and by the fact that several students used real-life examples as the basis of their answers. It is hypothesized that because of this, the comparison with course content or other students’ essays did not render meaningful results.

An important finding is that the system performs better when the course materials used for comparison are divided into sentences and not paragraphs. This is because the more separate passages of the course materials there are, the better is the distinction between the grade categories. This division also increases the sparseness of the training data that might have an effect on the distinction between grade categories.

Even though LSA seems to outperform the other two methods (PLSA and LDA), the differences, especially between LSA and PLSA, are small. PLSA also outperforms LDA in accuracy. Although PLSA and LDA have been shown to perform better than LSA in IR tasks (Hofmann, 2001; Blei et al., 2003), this was not the case in our experiment. One possible explanation for this difference might be the size of the collections that we used to train the model. LSA performs better with a small document collection. But there might be other explanations as well. While Hofmann (2001) and Blei et al. (2003) trained systems with collections of 1,000-3,000 documents in earlier studies, we used fewer than 150 documents. Since it is more likely from a practical point of view that relatively small collections of essays are graded, the results obtained from our experiments are valid in automatic essay grading context. It will be necessary to test these assumptions in future experiments if the size of document collections does indeed exert an effect on the performance of LSA, PLSA and LDA.

Although PLSA-C performs worse than PLSA, it is a useful method for removing the dimensionality selection phase in PLSA. This is caused by the fact that the differences between PLSA and PLSA-C are relatively small.

### Conclusion and Future Work

In this paper we have described an automated grading system that is based on comparisons between course materials and teacher-graded essays. We have also reported a set of experiments that used the system for comparing three dimensionality reduction methods, LSA, PLSA and LSA. We have also showed that the use of both course materials and human-graded essays makes grading more accurate than the k-NN-based grading method based on human-graded essays alone. The highest correlation between the grades assigned by the system and by a human grader (0.95) was achieved on the Software Engineering test set using PLSA for document comparisons.
The overall results show that our content-based grading model is acceptable in terms of costs (a low number of pre-graded essays is sufficient for training) and coachability (AEA is not based on surface measures). While our results show that all the dimensionality reduction methods that we have considered can provide an acceptable level of accuracy, it was rather surprising to observe that LSA slightly outperforms PLSA and LDA in the essay-grading domain. Because the probabilistic models of these two methods allow for other developments, they are worthy of further investigation. We observed, for instance, during the experiments with PLSA-C, that the combination of dimensions that lead to better results might be dependent on the features of the used document collection (such as the number of passages in comparison materials or the number of essays in training data). It is possible that the combination of dimensions can be optimized for each essay set by using such collection-specific features. This would obviate the need for using computationally demanding dimensionality selection methods such as those that are used for LSA.

The most important distinction between the three dimensionality reduction methods in the context of automatic essay grading is the way in which they support the defensibility of the grading system. PLSA and LDA offer a better means than LSA of giving students feedback about the essay by finding topics that are not covered by the essay writer because the topic distributions can be directly used as estimates of how well a specific topic has been covered. An essay-grading system that uses methods such as these can support a writer by offering keywords from each of the underlying topics that the essay has not addressed.

Defensibility is admittedly the weakest point of AEA (and other automatic grading systems). The next steps in research into AEA include increasing its defensibility by incorporating automatic feedback generation, writing style and plagiarism-detection modules into the system.

Acknowledgements

This work was supported by the Academy of Finland under grant no. 111350, “Automated Assessment Technologies for Free Text and Programming Assignments”. 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 Academy of Finland.

References


Blended learning: Using technology in and beyond the language classroom
(Book Review)

Reviewer:
Jesús García Laborda
Associate Professor in Applied Linguistics
Polytechnic University of Valencia
E. P. S. Gandía, Valencia, Spain
jgarcial@upvnet.upv.es
Tel: +34 666 28 45 57

Textbook Details:
Blended learning: Using technology in and beyond the language classroom
Pete Sharma & Barney Berrett
Macmillan
2007, 160 pages

Overview

Blended learning is the combination of delivery procedures in learning. In this century, blended learning has been defined as the combination of “face-to-face instruction with online learning using a learning management system” (Irons, Keel, & Bielema, 2002). However, with the new millennium, blended learning activities have been refocused and developed (Cooney, Gupton & O’Laughlin, 2000; Lim, 2002) although they had been present in the classroom almost from the beginning of the use of technology and computers in education. Today, when teachers are both aware of the importance of technology in the classroom and are also familiar with its use, blended learning has been approached with curiosity and interest (Donahue, 2001; St-amant, 2001). However, like any other activity in education, Blended learning becomes more attractive and increases its possible applications when teachers are well trained in its place in the classroom, and also when the teacher’s computer literacy improves (Kirschner, Clark and Sweller, 2006). In this sense, this book intends to cope with a necessity, especially of language teachers. Today nobody could deny that computer assisted language learning has been operational for over twenty years. However, although there is abundant literature that claims the benefits of blended learning because of the students’ satisfaction (Lai, Yeh, & Ho, 2005), many language teachers still fear approaching computers because they claim their lack of expertise especially when working with their students and that is the main reason why volumes that address computers abound in language education (a good range of them can be found in leading journals such as TESL EJ) (García Laborda & Magal Royo, 2007).

Overall, this volume is divided into nine chapters and also includes a not always accurate bibliography (check the mismatch in the sixth citation on page 8), two appendixes and an itemized index. The introductory chapter presents what blended learning is with a range of definitions but fails to give an in depth description of blended learning and its different models such as the skill-driven learning, the attitude-driven learning, and the competency-driven learning (for an extensive review of these model, please refer to www.Blended learning/Blended Learning Models.mht), the introduction is followed by a description of the book, a section on the benefits of using technology in the language classroom, factors that affect the implementation of blended learning and some general (but vague) guidelines of how to combine traditional teaching methods and teaching with technology. Although this section only intends to justify the use of technology in the classroom, from the authors’ individual perspective, these pages emphasize the need to incorporate communication technologies (and indeed the need of using technology in communication is emphasized throughout the book) in daily teaching. It also serves to motivate the readers who may hesitate to take a first step in the use of technologies in education and in face to face teacher-student interaction (p. 14).

The rest of the chapters in the book address the authors’ course contents on the same topics including:

- The web as a source: authentic materials, online activities, advanced search,
• Evaluating the English Language Teaching (ELT) materials found online,
• The use of online and electronic dictionaries,
• The use of whiteboard (this is an excellent chapter and I believe it is one of the few differences between this book and many others of its kid),
• The use of portable devices (a very well structured chapter and, in a way, very new even for those teachers who already use computers and ITC as a main component of their classes,
• Computer mediated communication (using both synchronic and asynchronic communication) such as chats, forums and so,
• Web resources such as podcasts,
• Suggestions about the future including the future use of web 2.0.

The last part of the book also includes two appendixes: a resource bank and some advice on how to get started, and the index and bibliography. Finally a well designed and welcome contribution are the three updates (between April 2007 and January 2008)[ http://www.macmillanenglish.com/Course.aspx?id=28614&producttypeid=28254] that add valuable information on dictionaries (April 2007), updates in methodology and virtual worlds (September 2007) and an extension on whiteboards (January 2008) (readers who have little idea of what a whiteboards will want to take a look to this information). This addition also proves the dynamic relation between paper and electronic books today.

Although the authors state that this book is aimed at all, teachers who have some experience will benefit the most because they may not fear the powerful advantages of computers and the net in teaching (especially in language teaching). To achieve this goal the authors describe the necessary resources and technology in each chapter. Usually it is quite clear but some language teachers may still be reluctant to use technologies in foreign language, and thus clearer instructions would have been advisable. I think the section which provides better information is that one in which Shama and Berret evaluate the individual use of technology provided in each chapter by pointing out its drawbacks and implications. I certainly agree with Brown (2007) that authors establish well the link between reality through the use of case studies and certainly their implicit belief that technology is for all the teachers with certain differences according to their own computer skills and that use leads to mastery. Along these thoughts, the writers emphasize the idea that teachers should overcome their fascination for the new usually related to new technologies and make connections between their teaching and learning technologies and what the ICT can offer. Needless to say, that as most other authors in the field whose work can be found in this and other journals (such as CALL, TESL EJ, Computers and Education, Language Learning and Technology, and many more), teachers should not be suspicious about the computer taking their teaching role but complemented and implementing their own teaching.

Overall, this is a well written volume for “false beginners” in ICT and (language) educators. Although the lack of a well defined framework is self-evident, the information provided in each individual chapter facilitates the inclusion of the activities presented in the book in the classroom. The approach is equally valid for language and general practitioners. I really missed some information for the most advanced advocates of using ICT in teaching and learning but the text quality outcomes its drawbacks. Thus, this book may well deserve a place in the library of many teachers around the world.

Bibliography


Semantic Web and Education
(Book Review)

Reviewer:
Fuhua Oscar Lin
Professor in Computing and Information Systems
Athabasca University, Canada
oscarl@athabascau.ca
Tel: 1-866-539-8664

Textbook Details:
Semantic Web and Education
Vladan Devedzic
Springer’s Integrated Series in Information Systems
ISBN: 0-387-35416-6

Because of the Web, we have Web-based Education. Putting Web based Education in the context of the Semantic Web creates the idea of the new generation WBE, or Semantic WBE (SWBE).

The Book “Semantic Web and Education” by Prof. Vladan Devedzic of at the University of Belgrade discusses how to use the results and technology of new fields of research and development, such as Semantic Web and Web intelligence, to make Web-based Education more effective and more appealing to learners, teachers, and authors alike.

The author integrates the examinations of learning-oriented topics such as learner modeling, collaborative learning, learning management, learning communities, ontological engineering of Web-based learning, and related topics with these technical topics.

It is a thorough and highly useful presentation on the confluence of the technical aspects of the Semantic Web and the field of Education or the art of teaching. The book will be of considerable interest to researchers and students in the field of Artificial Intelligence, Information Systems, and Education.

What impressed me first in Chapter 1 are the clear explanations of many commonly-used terms such as Web-based education, e-learning, distance education, virtual classrooms, and intelligent educational systems.

Chapter 2 introduces Semantic Web technologies and explains common prerequisites for creating intelligent Web-based Education systems and applications.

Chapter 3 explains the setting for SWBE replying on existing technologies and technological trends, and starting from the needs and perspectives of the major categories of actors in Web-based Education.

Chapter 4 discusses the architectural issues including architectural reference model, learning object structure and organization, current trends and technologies, and open learning environments.

Chapter 5 discusses the relationship between Learning Technology Standardization efforts and Semantic Web Based Education (SWBE), in particular Semantic Web Issues related to Learning Technology Standards. This chapter is very useful to me as the author pointed out some additional requirements that need to be addressed by standards and specifications. For example, the author mentioned that current standards are not published with formal semantic into existing standards and as a consequence they do not support reasoning and semantic search based on Learning Object metadata. Therefore we should introduce formal semantics into existing standards. However, it is easier said than done.

Chapter 6 deals with a central issue of SWBE --- personalization. Learner modeling enables personalization. The challenge is to capture useful information from the interactions by learners and SWBE systems as automatically as
possible, with minimum efforts from the learner. I enjoyed reading the case study TANGRAM, described in Section 3 of the chapter. It illustrates the process of ontological engineering of SWBE systems by explaining the design of the TANGRAM environment and by suggesting some more general engineering guidelines for building SWBE systems.

Chapter 7 further addresses the important topic --- Ontological engineering, which is a central process in developing any SWBE application or system. It is an excellent material to learn about the structure, context, usability, and technological background for using ontologies in education. This chapter introduces several new tools specific to SWBE that help designers and authors develop their applications and learning material. It also discusses ontology visualization techniques and tools, as well as automatic construction of educational ontologies.

Chapter 8 illustrates the use of SWBE principles, architectures, and technology in practical applications. It covers learning management, collaborative learning, and learning communities. Two newly emerged research topics, personalized educational services and learners’ personality, are also introduced in this chapter. I feel that the previous experience from architectural design of more traditional e-Learning systems, virtual classrooms, Web-based ITSs, and AEHSs are very useful to my current research, intelligence in adaptive learning systems.

While many of the related technologies and standards are still under development, Prof. Devedzic’s this book already offers both a broad conceptual introduction and lots of points to future application scenarios for researchers in academia and industry as well as for developers of Web-based educational systems. It is an extremely valuable addition to the AI and Software Engineering literature.
National Library of Virtual Manipulatives  
(Website Review)

**Reviewer:**  
Hosin Shirvani  
University of Texas-Pan American  
Department of Curriculum and Instruction  
Edinburg, Texas  78501  
hshirvani@utpa.edu

**Site URL**  
http://nlvm.usu.edu/

**Site Title**  
National Library of Virtual Manipulatives

**Objective of the Site**

It helps students understand mathematical concepts by using interactive manipulatives.

**Intended Audience**

It is mainly designed for students and the parents. However, teachers can use this site to teach their students mathematical concepts. University instructors can use this website in their classrooms teaching their pre-service teachers mathematical concepts by using interactive manipulatives.

**Domain Related Aspects**

This site is an educational site designed for teaching mathematics to elementary, middle and high school students. The site includes many types of manipulatives that teachers can use in motivating their students in learning mathematical concepts.

**Structure of the Site:**

The website’s mathematical concepts are divided to five areas: number sense and operations, algebra, geometry, measurement, and data analysis and probability. Each of this area is divided to four grade-level sections: Pre-K-2, 3-5, 6-8, and 9-12. Students can click on the content area they are interested.

**Usefulness and Richness of Each Topic**

I recommend this site to mathematics teachers, specially elementary mathematics teachers because this site contains many interactive mathematics activities which enhance student learning. The philosophy of this site is based on understanding of mathematics concepts rather than drilling and memorizing them. Many elementary mathematics teachers do not have manipulatives such as base ten blocks, tangram, geoboard, pattern blocks, attribute blocks, and others resources in their classrooms. They simply can get access to this Web in just a few minutes and click on these manipulatives designed at different grade levels. By using this Web, students are active participant in learning, for example to add 37+18, the web would display 3 longs, which represent 30, and seven small units, which represent the number even, and for 18, it will display one long and 8 units. The students should combines a bundle of 10 units
which will equal to one long, then this long will move to the other four longs. This shows students the reason that they need to add the number one to the ten-digit column.

**Connectivity**

It’s easy to connect to this Website as long as the user has a Java-supported browser. It works with both Apple or Microsoft-supported computers.

**Interface Related Aspects**

**Layout of the Website**
It has a well-designed layout, and it’s easy to move from link to another one.

**Site Structure**
It’s easy to read the site with no advertisements on it. It’s has a colorful design, which motivates children to use it.

**Navigation**

It is easy to navigate; it’s also fast. There is no downloading of any software, and the users need to click on the topic they are interested in.

**Search facility**

It’s easy to find any mathematical concept that the users are interested to learn. Just click on the proper math concepts and then click on the grade level. There is also a place that the user can click to get access to all listed manipulatives or concepts.

**Overall Issues**

The web was supported by National Science Foundation (NSF) and designed by Utah State University. There is an update for this website ever year.

**Other comments**

I have seen many mathematics websites designed for students. However, this site is the only one that encourages deep understanding of mathematical concepts with many activities for different concepts at different grade levels. Above all, It is free to the public.