Educational Technology & Society
An International Journal

Aims and Scope

Educational Technology & Society is a quarterly journal published in January, April, July and October. Educational Technology & Society seeks academic articles on the issues affecting the developers of educational systems and educators who implement and manage such systems. The articles should discuss the perspectives of both communities and their relation to each other:

- Educators aim to use technology to enhance individual learning as well as to achieve widespread education and expect the technology to blend with their individual approach to instruction. However, most educators are not fully aware of the benefits that may be obtained by proactively harnessing the available technologies and how they might be able to influence further developments through systematic feedback and suggestions.
- Educational system developers and artificial intelligence (AI) researchers are sometimes unaware of the needs and requirements of typical teachers, with a possible exception of those in the computer science domain. In transferring the notion of a ‘user’ from the human-computer interaction studies and assigning it to the ‘student’, the educator’s role as the ‘implementer/manager/user’ of the technology has been forgotten.

The aim of the journal is to help them better understand each other's role in the overall process of education and how they may support each other. The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

The scope of the journal is broad. Following list of topics is considered to be within the scope of the journal:

Editors

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

Editors’ Advisors

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

Editorial Assistant

Barbara Adamski, Athabasca University, Canada.

Associate editors

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

Advisory board

Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Ruddy Lelouche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Richiroy Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, National Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan.

Assistant Editors

Yuan-Hsuan (Karen) Lee, National Chiao Tung University, Taiwan; Weichieh Fang, National Sun Yat-sen University, Taiwan.

Executive peer-reviewers

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

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

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

Abstracting and Indexing

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

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

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

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

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

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

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

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

### Special issue articles

- Guest Editorial - Navigating in the Digital Era: Digital Literacy: Socio-Cultural and Educational Aspects  
  *Yoram Eshet-Alkalai and Oren Soffer*  
  1-1

- The Unexpected Connection: Serendipity and Human Mediation in Networked Learning  
  *Rita Kop*  
  2-11

- How Do Personality, Synchronous Media, and Discussion Topic Affect Participation?  
  *Ina Blau and Azy Barak*  
  12-24

- Internet Abuse among Teenagers and Its Relations to Internet Usage Patterns and Demographics  
  *Suzan Lema Gencer and Mustafa Koc*  
  25-36

- Digital Assessment Literacy - the Core Role of the Teacher in a Digital Environment  
  *Liat Eyal*  
  37-49

### Full length articles

- Integration of Multimedia Courseware into ESP Instruction for Technological Purposes in Higher Technical Education  
  *Shu-Chiao Tsai*  
  50-61

- An Empathic Avatar in a Computer-Aided Learning Program to Encourage and Persuade Learners  
  *Gwo-Dong Chen, Jih-Hsien Lee, Chin-Yeh Wang, Po-Yao Chao, Liang-Yi Li and Tzung-Yi Lee*  
  62-72

- Evaluating Knowledge Structure-based Adaptive Testing Algorithms and System Development  
  *Huey-Min Wu, Bor-Chen Kuo and Jinn-Min Yang*  
  73-88

- Exploring Learner Attitudes toward Web-based Recommendation Learning Service System for Interdisciplinary Applications  
  *Hong-Ren Chen and Jhen-Gang Huang*  
  89-100

- Passively Classifying Student Mood and Performance within Intelligent Tutors  
  *Robert A. Sottilare and Michael Proctor*  
  101-114

- Integrating Algorithm Visualization Video into a First-Year Algorithm and Data Structure Course  
  *Pilu Crescenzi, Alessio Malizia, M. Cecilia Verri, Paloma Diaz and Ignacio Aedo*  
  115-124

- An Assessment of E-training Effectiveness in Multinational Companies in Malaysia  
  *Thurasamy Ramayah, Noor Hazlina Ahmad and Tan Say Hong*  
  125-137

- Supporting Teachers in Designing CSCL Activities: A Case Study of Principle-based Pedagogical Patterns in Networked Second Language Classrooms  
  *Yun Wen, Chee-Kit Looi and Wenli Chen*  
  138-153

- A Clustering Methodology of Web Log Data for Learning Management Systems  
  *Stavros Valsamidis, Sotirios Kontogiannis, Ioannis Kazanidis, Theodosios Theodosiou and Alexandros Karakos*  
  154-167

- Blogging for Information Management, Learning, and Social Support during Internship  
  *Samuel K. W. Chu, Alvin C. M. Kwan and Peter Warning*  
  168-178

- Using Mobile Learning to Improve the Reflection: A Case Study of Traffic Violation  
  *Yu-Feng Lan and Shin-Ming Huang*  
  179-193

- The Comparison of the Opinions of the University Students on the Usage of Blog and Wiki for Their Courses  
  *Ummuhan Avci and Petek Askar*  
  194-205

- Educational Affordances of a Ubiquitous Learning Environment in a Natural Science Course  
  *Tan-Hsu Tan, Min-Sheng Lin, Yu-Ling Chu and Tsung-Yu Liu*  
  206-219
Contextualizing a MALL: Practice Design and Evaluation
Yaming Tai 220-230

Reducing the Impact of Inappropriate Items on Reviewable Computerized Adaptive Testing
Yung-Chin Yen, Rong-Guey Ho, Wen-Wei Liao and Li-Ju Chen 231-243

Modelling Digital Natives' International Collaboration: Finnish-Korean Experiences of Environmental Education
Irja Leppisaari and Okhwa Lee 244-256

Designing Hypercontextualized Games: A Case Study with LieksaMyst
Carolina Islas Sedano, Erkki Sutinen, Mikko Vinni and Teemu H. Laine 257-270

Understanding Social Nature of an Online Community of Practice for Learning to Teach
I-Chun Tsai 271-285

Cyber Asynchronous versus Blended Cyber Approach in Distance English Learning
Zi-Gang Ge 286-297

Adaptively Ubiquitous Learning in Campus Math Path
Shu-Chuan Shih, Bor-Chen Kuo and Yu-Lung Liu 298-308

The Development of Epistemic Relativism versus Social Relativism via Online Peer Assessment, and their Relations with Epistemological Beliefs and Internet Self-efficacy
Chin-Chung Tsai 309-316

Influence of Game Quests on Pupils' Enjoyment and Goal-pursuing in Math Learning

The Effect of Principals' Technological Leadership on Teachers' Technological Literacy and Teaching Effectiveness in Taiwanese Elementary Schools
I-Hua Chang 328-340

Instructor Learning Styles as Indicators of Online Faculty Satisfaction
Ryan McLawhon and Marc Cutright 341-353

Yu-Chen Hsu, Hsin Ning Jessie Ho, Chin-Chung Tsai, Gwo-Jen Hwang, Hui-Chun Chu, Chin-Yeh Wang and Nian-Shing Chen 354-370
Guest Editorial - Navigating in the Digital Era: Digital Literacy: Socio-Cultural and Educational Aspects

Yoram Eshet-Alkalai and Oren Soffer
The Open University of Israel, Israel // yorames@openu.ac.il // orenso@openu.ac.il

In recent years, digital technologies (social media, multimedia and communication technologies) have penetrated almost every aspect of our lives. They’ve become an essential element in learning, interpersonal communication, organizational work, and our leisure time. Extensive research has recently undertaken to assess how digital technologies—such as e-book readers, social media, mobile technologies and multi-player online games—are actually altering the traditional modes of learning, knowledge construction, cognitive and social skills, information distribution and consumption, as well as social participation and relationships.

Ever-expanding and changing digital technologies confront users with the need to master a wide range of technological, cognitive and social competences—collectively termed "Digital Literacy." Users thus must become "digitally literate" in order to cope effectively with the complex sociological, cognitive and pedagogical challenges these technologies pose. These skills include, for example, the ability to operate computers and navigate the net effectively, to cope with large volumes of information, to evaluate the reliability of information, and to critically assess what seem to be natural (and not ideologically biased) technological tools. In a different way from the spirit of modern print, learners construct and consume knowledge in non-linear environments. They need to learn, collaborate and solve problems effectively in virtual (non face-to-face) learning environments, and to communicate effectively in technology-mediated social participation environments.

It is important to note: digital literacy, then, is not limited simply to computer and Internet operation and orientation. It also relates to a variety of epistemological and ethical issues arise due to the unique characteristics of digital technologies and that are often overlapped with trends related to the post-modern and post-structural era. These include questions regarding the authority of knowledge, intellectual property and ownership, copyright, authenticity and plagiarism. Furthermore, issues such as self-representation, virtual group dynamics, and on-line addiction also arise. In this sense, digital literacy allows learners and users to navigate the changing socio-technological and philosophical-epistemological environments, where the conventions regarding preferred and prohibited learning methods are constantly challenged—both by social and philosophical trends and by communication and learning technologies.

This special theme issue of the journal focuses on the various aspects of digital literacy. It explores the expression of digital literacy and digital literacy skills in fields such as learning, teaching, media affordances, and sociology. It does so by discussing the challenges and obstacles to digital literacy's optimal effective and efficient utilization of innovative technologies.

The four articles in this special issue vary in their topics of study and research methods, ranging from an empirical quantitative study to theoretical papers. Rita Kop’s article relates to learning epistemology on the Internet, as well as to user- awareness of the non-natural nature of the technology of search engines. Kop stresses the restrictions and limitations that these engines impose on the users while browsing information through the net. According to Kop, this limited search environment—which is radically different from the popular image of a democratic and free Internet—stresses the need for human mediation of information in the learning process. The next two articles in the journal demonstrate behavioural research perceptions. Ina Blau and Azi Barak explore the variables that affect people’s participation and contribution in online group discussions. Among other things, the article compares participation rates and the characteristics of online textual or audio chats in different social contexts. The paper by Suzan Lema Gencer and Mustafa Koc deals with dysfunctional digital literacy. Their article examines Internet abuse among Turkish teenagers: drawing upon a wide and varied student sample, the study assesses the role of social and academic backgrounds in explaining such abuse. And finally, in the concluding article, Liat Eyal discusses the role of the teacher in the digital era, with a special focus on digital assessments.
The Unexpected Connection: Serendipity and Human Mediation in Networked Learning

Rita Kop
National Research Council of Canada, Canada // frederika.kop@nrc.gc.ca

ABSTRACT
Major changes on the Web in recent years have contributed to an abundance of information for people to harness in their learning. Emerging technologies have instigated the need for critical literacies to support learners on open online networks in the mastering of critical information gathering during their learning journeys. This paper will argue that people will have to adapt to using information in a new way and will advocate the movement by learners into and inside information streams on open online networks. Their own control and aggregation of information, preferably through human mediation, should provide information not only relevant to their learning, but also slightly unexpected. We will highlight why this serendipity is important in a learning context and also take three emerging technologies under the loupe; recommenders, RSS and micro-bloggers, and their effectiveness in supporting serendipitous learning on open online networks.

Keywords
Information literacy, Recommender, RSS, Twitter, Serendipity, Networked learning

Learning in a new environment characterized by complexity
The context of learning has changed dramatically in recent years. In the past the learning experience was determined by three dimensions: the learner, the educator, and the content, but a fourth aspect has increased in prominence in theories and practice of learning: the learning context (Kop, 2011). Bouchard would even go so far as stating that: “For the first time we are understanding the act of learning as a response to changes in the learning environment, rather than as an adaptation to a predetermined learning system.” (Bouchard, 2011a, p. 290). Educators are re-examining their practice in the light of emerging technologies that have provided learners with an explosion of resources. Only ten years ago the book was the major source of information, but the Web is currently developing into a mesh of interlinked resources that can be accessed through communication and collaboration with others. This development increases the volume of information available, adds to the complexity of the educational and learning landscape, but it also invigorates it.

Barnett highlighted that we now live in a world characterized by “super-complexity,” uncertainty and change: “Work, communication, identity, self, knowing and even life: the meaning of fundamental concepts are no longer clear in a world of change” (Barnett, 2002, p. 9). He referred not only to technological change in society and education, but it remains that technological change has undoubtedly created a more complex learning environment than we were used to not so long ago. Learning is at the heart of personal change and transformation and Folke (2010), while discussing “resilience,” reminds us that people have shown to be fairly resilient in adapting to the changed environmental realities of the new information landscape. However, the accelerated pace of technological change has also shown to be problematic and might require professional adaptation from educators and from learners alike. This paper will reflect on changes in information gathering in relation to learning, and the role that recommender systems, RSS and micro-blogging might play in facilitating improvements and innovation in the aggregation and validation of information.

The Web: The nature of the network—Authority and Inequality on the network
We have seen a substantial growth and development of the Web over the past several years. The Web only ten years ago mainly consisted of fairly static content and basic sharing opportunities. This has now changed with the addition of social media, the cloud, the use of Web semantics and mobile and wireless technologies. Content is increasingly produced by users and there has been an exponential growth in the use of social media such as blogs (Smith, 2008), and video-sharing sites such as YouTube. More than 70% of the Web is now user-created (Bloch, 2011) and distributed via personal presence sites such as Facebook and YouTube in addition to micro blogging sites such as Twitter. All this data is stored, used and re-used in different ways and mashed up into new information.
The Web is portrayed as a democratic network on which peer to peer interaction might lead to a creative explosion and participative culture of activity. It should be noted, however, that most new, successful grass-roots developments on the Web have been commercialized and integrated in the corporate world. Increasingly concern is being raised about the influence of commerce on the Web. Lanier (2010) and Mejias (2009) emphasize the high level of influence by a low number of companies, such as Google and Facebook. Market forces seem to slowly but steadily influence and take control of the new tools. The users are no longer the customers important to the directors of these applications, but their “social graph” has become the product that is “sold” to the advertisers who bring in revenue. The freedom and creative potential of the Web for all and the information available seems to be progressively more influenced by other interests.

This is exemplified by the development of Google. Google accounts for 72.15% of all searches in the USA (Experian Hitwise, 2010) and 90% in the European Union (White & Campbell, 2010), which makes Google’s behavior and its integrity in relation to the access it provides to information crucial to networked learning. It might be a little naïve to expect a search engine to behave in the best possible interest of the searcher, after all as Grimmelmann explains:

> Search engines are attention lenses; they bring the online world into focus. They can redirect, reveal, magnify, and distort. They have immense power to help and to hide. We use them, to some extent, always at our own peril. And out of the many ways that search engines can cause harm, the thorniest problems of all stem from their ranking decisions. (Grimmelmann, 2010, p. 435)

We will return to the ranking and recommender systems later on in this paper as they are currently being developed to facilitate learning. As educators it seems to become vital to assess the best possible ways to collect information and voices can be heard advocating a publicly funded search engine, rather than one controlled by commerce to avoid bias in the searches we carry out (White, 2010, Goldman, 2010). As White argued compellingly:

> The danger of allowing an advertising company to control the index of human knowledge is too obvious to ignore. The universal index is the shared heritage of humanity. It ought to be owned by us all. No corporation or nation has the right to privatize the index, commercialize the index, censor what they do not like or auction search ranking to the highest bidder. We have public libraries. We need a public search engine. (White, 2010, p. 1)

A body of research is emerging that investigates the Web network itself. Barabasi looked at the mathematics of the Internet and Web as networks and found that they do not perform as “random” networks, but as “scale-free” networks. The difference would be ruled by two characteristics: “growth” and “preferential attachment”, showing that this type of network grows “one node at a time” and that a node must “choose” to what other nodes it will connect. However, at the same time, the more connections a node has, the more likely it is that other nodes will attach to it (Barabasi, 2003, p. 86). This phenomenon creates “hubs”, defined as nodes that have a large number of links directed to them. The early adopter nodes attract a multitude of other nodes, but it is not necessarily their value that is attractive, more likely their popularity and attraction to others. This means that there are clearly power-relations on the network and Barabasi’s research shows that networks are not neutral, which is also emphasized by Bouchard (2011a). Barabasi found in his research that participants on networks are not only selective, but that the nature of networks and the “power curve” prevents network “surfers” from having access to all information at the same level.

Even though it seems on the surface that people have access to any piece of information and resource they would like on the Web, in reality this access is restricted by the structure of the Web and by the ranking of information by search engines (Grimmelmann, 2010; Goldman, 2010). It is clear that the ability to understand the intricacies of the network is required by people to negotiate this structure when learning on these vast disparate information networks. As self-directed learning becomes more prevalent, the need for high levels of critical capabilities, in addition to knowledge of the sub-sytems of the Web, are important to be able to access information and resources that are relevant and required to advance learning. It should be questioned, however, whether all adult learners are able to do so without help from knowledgeable others (Kop & Bouchard, 2011). My own research (Kop, 2010) shows that it is the presence and involvement of (knowledgeable) others in an environment characterized by many technological variables and contexts that help learners to make sense of the multitude of resources offered on the Web. In the absence of adult educators, it is imperative to harness the changing affordances and potential of new technologies to this extent.
Human mediation and information flows in open networked learning

People’s information habits have “deeply ingrained habitual patterns” (Fischer & Naumer, 2006, p. 2). It seems that people will first and foremost find information from people with whom they have a strong relationship, which are usually found in their circle of family, close friends and their local communities in places such as doctors’ surgeries and libraries. The Web is supplementing these “information grounds” and is also creating new structures for obtaining information.

CIBER (2008) highlighted how people acquire information and how information behavior has changed over time. They found that “power-browsing,” the clicking of hyperlinks and the skimming of web pages, has replaced traditional chronological reading and longer term critical thinking, and that advanced information searching is lacking. Their research showed low levels of information literacy, in the form of validating information and sources. Information literacy is acquired at a young age and “information behavior” is a developmental process at a deep level that will be very difficult to advance substantially later in life, for instance on a course at university (CIBER, 2008). The abundance of information on the Web has raised concerns about the feasibility for individuals to critically analyze all that is available to ensure reliability and validity and to manage the vast streams of information now available. As Hagel (2006) observed, the more information is available, the less time we have available to go into any depth when analyzing the information as the overpowering amount of information we have to deal with reduces our attention.

We therefore seek people who distribute the information we are interested in; they will get the attention, but we trust them to deliver the information relevant to us. Human behavior is influenced by people in their direct environment, with whom they have strong ties, but also by people with whom they have weaker ties, such as on social networks (Granovetter, 1973) and studies related to information flow between mass media and individuals have been conducted for many years. Katz and Lazersonfeld (1955) for instance developed the two-step flow theory related to communication and their research indicated that mass media information is channeled to “masses” through human intermediaries who would have a high level of access to and understanding of the media. With the inception of the Web this form of ‘one to many’ communication, perhaps via an intermediary, has been supplemented by a form of “many to one to many” communication on digital social networks, where a commercial company has taken over the role of intermediary who controls the flow of information (Mejias, 2009). In networked learning information brokers might be hubs on networks, who recommend information, and to whom people are attracted because of their reputation. Studies using social network analysis show that opinions and beliefs are influenced by these “committed agents,” who are not easily influenced themselves, but spread their message and information until a ‘tipping point’ is reached at which their “followers” adopt their point of view (Granovetter, 1978; Xie et al., 2011, p. 1). People are clearly influenced in their information behavior by these human conduits. The problem of course is that the reputation of these intermediaries is acquired not only because they have proven to supply interesting information, but also through the process described by Barabasi as “preferential attachment.” Information brokers become powerful distributors on the network and make choices about the information and resources that are filtered to others (Boyd, 2010). One could argue that it would be more appropriate for this power to reside with the creators of the information, or with the information gatherers themselves as the broker does not have a responsibility to validate the information before distributing it as educators do. The “many to one to many” form might be well-suited to networked learning as long as the commercial intermediary does not have too prominent a role in influencing the information flow. This would mean that the social influence would not come from one or two prominent information brokers, but from a multiple of people to increase the reliability of the information. Social influence has always been important in the access of information on learning networks, but has evolved. Educators are no longer the main information providers with a responsibility to safeguard its quality and who serve learners with a mixture of information not only giving learners what might be exactly relevant in a search, but also the opposite, or perhaps the unexpected to make particular points or to engage learners in a thinking process. We would argue that as educators might not be available during networked learning, it is this mixed supply of information provided through the mediation of people that could advance self-directed networked learning.

Distinguishing the chaff from the wheat

With the emergence of social media web users themselves can now be in control of their information aggregation; they don’t necessarily have to use information brokers or educators to validate their information. Networked learning
might take place outside the boundaries of educational institutions and one of the major tasks for any autonomous learner on an open online network is to identify useful elements within the overwhelming volume of unsorted information. Search engines do some sorting (or perhaps sorting to reflect commercial or other interests), and learners need to make connections between these resources in order to engage in thinking and learning, and possibly produce something new to share with others.

As the number of information sources has increased dramatically and information has also become more distributed, the need to work with information in a different way has become indispensable in order to evaluate its quality and maintain some coherence. The openness of the Web has made that people have access to growing numbers of data, not only comprised of basic pieces of writing, but also social interactions and cultural artifacts. As mentioned before, the Web is not a power free and hierarchy free environment, and people have already made a start with organizing their own streams of information and activities using information hubs, human filters, lists, tags and #tags.

An information folksonomy has emerged, a classification system not based on library or academic classification systems, but on the ordering of information sources by people, for instance in the form of key words such as tags and #tags as identifiers. This means that information that users value can be organized and stored in databases and retrieved in the format required. Boyd, however, would like this to go one step further, and believes that people should have access to tools

that allow them to get in flow, that allow them to live inside information structures wherever they are and whatever they are doing. They need tools that allow them to grab what they want and to stay peripherally aware without feeling overwhelmed. (Boyd, 2010, p. 2)

But how can this be done without losing control? And how might these information streams actually promote learning?

The semantics of the web and the storage of information in databases are increasingly important in the identification and categorization of information. Google for instance uses their algorithms to crawl the Web and find relevant information related to the search key words. This use of semantics and database storage makes it possible to personalize information. Computer scientists are currently engaged in developing sophisticated tools to help learners make sense of information. They are developing recommender and ranking systems that work in a similar fashion as the systems that internet firms such as Amazon apply to web-searching and purchasing behavior, and the behavior of “friends” (friends in the sense of social networking friends, or customers who bought the same book before) (Andre, Schraefel, Teevan & Dumais, 2009). One would of course question who decides on the content and the values within the algorithms and also what would be the conceptual drivers behind the algorithms—and would they be any use for learning if any form of human mediation in searching and information provision is lacking?

Learning recommendation based on relevance or serendipity?

An interesting discussion is currently taking place in the literature about the possibility and even the desirability of algorithms, being mathematical formulas, to make decisions about the resources that will be recommended to Web users. Of course to a certain extent the search engine algorithms are influenced by human editors (Goldman, 2010), but an important component to a search is trust; and could we ever trust a machine, even though it is tweaked by humans, to find really useful information for us? In the past we might trust educators to make decisions about the resources that could enhance our study, or we would go to the library and browse the shelves ourselves or use library services to find what we needed, but in the extended open networked environment in which people now learn, those behaviors have been shrinking. This was highlighted in research by Pardo and Kloos (2011) who used a virtual computer to analyze the information behavior of their students and found that students only used the university resources for 28.51% of their study. For the other 71.49% they moved outside the institutional boundaries and searched the Web. Our information supply is evolving into a mesh of interlinked online resources that could be powerful in learning if we are able to find what is relevant to the purpose of our search.

Learners take advantage of commercial search engines, based on algorithms that make decisions about information they receive on a daily basis. Search engines, however, don't necessarily cater to advanced intellectual inquiry as their top search results merely reflect the general information needs of the population as a whole by bringing up
relevant information based on some key words. Google and Facebook algorithms provide us “with the information that they think we want to see, rather than all we can—and should. . . . The way algorithms work means that the focus is on what we click on most often, rather than providing us with a ‘balanced information diet’ that also include things that are uncomfortable and challenging and that include other points of view” (Zetter, 2011, p. 1).

Haraclitus already highlighted in 500BC that “the unexpected connection is more powerful than one that is obvious” (Hurson, 2007) and Gritton (2007) sets out a number of other purposes that people might have for particular searches to complement the directive one facilitated by search engines. These purposes would determine the best tools and search strategies to access the information. A search might be directive, to find a particular piece of information, or it might be capricious in nature, free flowing and characterized by a random move from link to link, or it could be semi-structured looking for inspiration. Another possible option could be characterized by browsing on sites that one might expect to contain interesting information.

Algorithm-based search engines and recommenders are very good at aiding in directive searches, but they are not so good at replicating serendipity; the chance of finding a gem of information, unrelated to a focused search, more as a by-product, that stimulates our creativity and thinking to arrive at a particular insight (Falconer, 2010; Andre et al., 2009). Andre et al. (2009) researched serendipity to try to understand how it could be introduced in recommender systems and posited that serendipity consists of two components: the finding of some unexpected, surprising, interesting information, and then by making connections to what is already known, perhaps in a particular domain, a creative insight might follow. They argue that the first part, the inclusion of the finding of unexpected information, might be facilitated through recommender systems, but the second part, the insight that might follow, which is related to learning, is much harder to achieve. As highlighted by Falconer:

The interlinked nature allows for increased opportunities for serendipity, but any automatic “enhancement” of what should be a wander through the park sniffing whichever flower takes our fancy, being told by any external agent what a flower might smell like. No. That’s no longer serendipity, that’s an enhanced search tool or an augmented results parser or whatever you might like to call it. To call it serendipity implies that you believe that some portion of your mind’s complexity, your memories and all the intuitive associations that go to make up you can be split off and embodied in a piece of maths.

Take a good hard look at what cutting edge neuroscience is actually telling us about our minds and how well we understand them. You’ll see that we are nowhere near being able to claim to be able to simulate serendipity, even if it were desirable to do so. (Falconer, 2010, p. 4)

Of course we have to balance the abundance of information with the ability to sift through it and find the best pieces for our learning needs. Algorithm-driven computer applications have so far not been able to automate serendipity, but only managed to go as far as suggesting content that “may be perceived to be serendipitous” (Andre, 2009, p. 309). Some authors highlight that serendipity is an important aspect in searching for information and in the creation of knowledge (Foster & Ford, 2003). Moreover, Gritton (2007, p. 6) argues that, “Serendipitous browsing does however have the potential to reveal connections between ideas that may otherwise go unnoticed, to stimulate ‘out-of-the-box’ thinking, and to challenge our mental models so that new learning can take place. In this regard, serendipity, free association and aimless browsing can lead to serendipitous learning” and is well worth pursuing. Furthermore, serendipitous learning is associated with “gaining new insights, discovering interesting aspects and recognizing new relations, which occur by chance or as by-product of other activities” and are related to people’s interest, prior knowledge and the setting of learning goals (Buchem, 2010, p. 1). If in an educational, or in a “networked learning” context outside formal education, critical and creative thinking is valued and seen to be important to advance learning, serendipity seems an important concept to strive for. After all, there are numerous examples in the literature to show that important discoveries and insights were facilitated by serendipity (Andre et al., 2009). But how could serendipity be stimulated in an autonomous learning environment?

Perhaps the greatest challenge in conducting a fruitful serendipitous investigation would be a change of our search strategies from looking something up and relying on brokers and search engine algorithms to filter our search results, to facilitating more randomness in our information stream. We must take control and find ways of incorporating web-searching into our thinking and reflection processes and pulling these processes into our own technological system that streams our information; the information on which we have decided ourselves, that is related to our own personal context, and from which we can pick and choose ourselves; an unfiltered but manageable store of resources.
Bouchard (2011) believes that even this is not enough, which there is something vital missing from this picture. To be an interesting component in a learning context and for a learner to use the information to create knowledge and to advance his or her learning, it would be desirable for the information not only to be filtered by ourselves, but to also be validated by other human beings:

For the first time, we as human beings are doing something that computers do not, namely recognizing each other as participants in ongoing conversations about the fluid nature of knowledge, and then attributing value to that knowledge as an evolving, changing thing. (Bouchard, 2011a, p. 294)

That is what current social media add to the earlier Web 1.0 developments: opportunities to share and communicate. Not just receiving information from one broker, or the mass media, but from a multitude of people on our network. The challenge would be to manage this stream effectively without being overwhelmed by the volume. New emerging collaborative services, such as micro-blogging tool Twitter and curating tool Scoop-It, facilitate networking, communication and sharing options with others, while information syndication tools, such as RSS aggregators and RSS readers, facilitate advanced search options that could aid in this development. One of the important factors of using these RSS and micro-blogging strategies would be that learners have full control over their information, resources and contacts, and over the access to their information.

**RSS—the simple aggregation of information feeds**

Rich Site Summary, or more popularly known Really Simple Syndication (RSS) was originally designed in 1997 by the web browser Netscape to customize web pages and to list the changing content on News websites (Downes, 2004). At the time it did not take off, but slowly and surely RSS has grown in popularity. So what does RSS do exactly? It delivers the news, and new developments that interest the user, directly and dynamically to her desktop. If creators of web resources include an RSS identifier to their site, people can subscribe to the stream of content produced on the site or blog, and RSS will then identify new items and display a “headline” from this content in an RSS reader through a regular synchronizing process, which means that the content is displayed together with other chosen feeds to produce a continuous stream of customized information and resources. This is a very convenient and efficient way of browsing through a multitude of new items and of course, the user controls the system; she can add or delete any stream that not quite matches her expectations, or even give a particular rating to some especially favorite ones and store it in one of her favorite bookmarking sites.

**Micro-blogging tools for the sharing and amplification of information**

The increase in popularity of micro blogging tool Twitter means that RSS links and feeds found through an RSS reader are being “amplified” and redistributed to contacts and followers that people have on their network in the Twitter communication environment. How does this work?

Twitter, a social media application that was first established in 2006, allows users to post messages of no more than 140 characters long and send these to people who are following them. People can follow others who they believe have something interesting to share. Users have started particular practices of including identifier codes in their messages, in the form of @ and #. This ensures that a particular person is identified in a particular message related to him in the case of the @, and that people can identify and search all messages related to a certain subject by using a keyword in combination with a #, producing a hash-tag, and can store this search in the form of a list. In addition, it is possible for users to amplify the same message they received from someone else, and pass it on to his own network of followers through the click of one button, which is called re-tweeting (Boyd et al., 2010). Microblogging messages have evolved to not just include a quick message about a certain topic, but to also contain links to web sites or blog posts, in addition to displaying these items in daily newspapers, which are Twitter aggregators that draw content from a combination of messages and multimedia from different sites. Educators are currently also experimenting with curating tools such as Scoop-It, which provide opportunities to curate messages and information on a particular topic, and add a longer comment than is possible on Twitter.

Of course carrying out a search is a different activity with a different purpose to the aggregation of information through RSS and microblogging or curation, and each of these has its own place in supporting the information needs
of learners. A search facilitated through an algorithm-based search engine provides fast direct answers to a query, while the user-controlled collection of information, mediated by human beings, might add to the validation of the information and to the level of serendipity.

**How could these developments affect information behavior in networked learning?**

RSS and Twitter offer advanced options for networking, for receiving and sharing information, and it seems that we are in a new era, where on the one hand we reach back to a form of village gossip communication through Twitter (although as was shown during the recent uprisings in Arab nations we can now do this on a global scale), while on the other hand we seem to link this “chattering” with a highly sophisticated form of personalized information aggregation and distribution that makes this combination interesting for learning. In the words of Bouchard:

> We are confronted with what is missing from this picture, namely the negotiated construction of knowledge. This is perhaps the most intriguing development resulting from the advent of the network age, although it is more reminiscent of small village cracker-barrel exchanges than futuristic networks: humans need to agree on stuff, but before they can do that, they need to talk about it. (Bouchard, 2011a, p. 294)

Our human social nature makes that we communicate about what keeps us engaged, and in turn we reflect on activities and information and make connections with what is already known. In the process we validate our information and knowledge. The way we can aggregate RSS streams and pass links on to others through tweets and re-tweets means that we can now give ourselves and others access to a serendipitous stream of information. People we follow will send us pieces of information that they find interesting, while re-tweets will give us access to an additional network of people, perhaps with weaker ties to us than the people we follow, and consequently with a different angle to information than our own network. Moreover, the use of #tags to form networks and the incorporation of these in our messages gives us access to an even wider network of informants and information to raise the potential for serendipity. This information can then be stored for future use in social bookmarking applications, such as Delicious and Diigo, which in their turn also offer search and share facilities.

We had the opportunity to test the extent of the influence of human mediation on information gathering in networked learning through research on a Massive Open Online Course in the Fall of 2010. The course was not organized by one institution that distributed content, but resources and the learning environment were distributed over the Web on blogs and in social networks. RSS and Twitter were used by facilitators and participants to aggregate information related to the subject of the course, Personal Learning Environments, Networks and Knowledge, and as options for communication. The 1641 participants used Twitter extensively, 3402 messages in total, and it was one of the participants communications tools of choice, which increased in use, and was highlighted by participants as important in receiving new and unexpected information. Figure 1 shows a visualization of the PLENK2010 Twitter network and the connection between different participants, while Figure 2 shows how PLENK2010 participants were
involved in other Twitter networks through the use of hash-tags. They highlight that not everyone was connected, represented in the outlying participants in Figure 1, but that some were highly connected in their extensive use of hash-tag networks. This proved to increase the streaming of serendipitous information onto the PLENK2010 network from other networks. The networks which contacts were engaged in were also relevant to the information aggregator, but relevant one step removed, with the potential of “unexpected relevance,” as Jarvis calls serendipity (Jarvis, 2010).

**Increasing the level of serendipity on the learning network**

These are only some basic indications of the positive effect of micro-blogging tool Twitter on the level of serendipity in open networked learning and that serendipity might be increased through communications with other people. If we were to define what would heighten a “degree of serendipity” on a learning network, it would be using the following factors: 1. The level of control over the information gathering process. 2. The involvement of people. Not only the number of contacts would be important, but also the degree of separation of the contacts and networks from the information gatherer, as it seems a higher level of serendipity is achieved if the information provider is somewhat removed from the information collector, while a higher level of relevance is realized from contacts closer to the collector. 3. The aggregation of information feeds. Not only is the number of feeds important to achieve serendipity, but also the variation of these feeds.

One would expect that the higher the level of control, and the higher the level of micro-blogging re-tweets, #tags networks involved in, and the spread of RSS feeds would be, the higher the likelihood of unexpected and challenging pieces of information and serendipity in the aggregator’s information stream. This level of unexpectedness would also depend on the right level of distance between the aggregator and his contacts or feed/link providers. If the distance is too great, the information becomes too random and could become irrelevant, but if it is very close the level of unexpectedness shrinks. Especially receiving re-tweets would heighten serendipity as these tweets would be provided by contacts of contacts, so still be reasonably close to the aggregator. In the absence of an educator we would welcome a high level of serendipity in the networked learning environment as it would add to the level of critical reflection and analysis and depth of learning.

**Conclusion**

This paper has highlighted some challenges and opportunities for learners in dealing with the abundance of information available at a time when new technologies are emerging and shaping the new information landscape. The role of the educator might slowly but surely change, from provider of information to provider of guidance about technology used to aggregate information in the most suitable fashion to advance learning. Technologies are now available to empower the learner to take control over the information stream him or herself and start to be proactive in the quest for valuable information. In the words of Boyd: “As we continue to move from a broadcast model of information to a networked one, we will continue to see a reworking of the information landscape. Some of what is unfolding is exciting; some is terrifying” (Boyd, 2010, p. 2), but it is clear that to make the most of emerging technologies learners will have to take an active role in shaping their own learning environment and in controlling information flows and communication tools.

An area for future inquiry would be to research more in-depth how emerging technologies can help people with the process of shaping their kaleidoscope of information most effectively. We need to know how people position themselves at the center of this process and how they can ensure that it provides complex, colorful and shifting patterns that are shaped by human connections and interactions, in order to at times surprise and challenge them. Empirical research to further test a ‘serendipity index’ based on micro-blogging features and RSS feeds might help in predicting the level and increase of serendipity in the information stream.

It will also be important to research how levels of serendipity can be heightened in information gathering and especially if this will at all be possible by automated means through recommender systems. Even though attempts are being made to include serendipity in these new technologies, it seems that this can only be achieved at a very rudimentary level and that recommenders currently mainly provide information relevant to a search term used. The use of RSS and micro-blogging tools, such as Twitter, seem to be more promising in dealing with high volumes of
information as human mediation means that a higher level of serendipitous information is being collected than would be possible in an automated environment.

References


Granovetter, M. S. (1973). The strength of weak ties. The American Journal of Sociology. 78(6), 1360–1380.


How Do Personality, Synchronous Media, and Discussion Topic Affect Participation?

Ina Blau and Azy Barak

Department of Education & Psychology, Research Center for Innovation in Learning Technologies, Open University of Israel, Israel // 1Department of Counseling & Human Development, Department of Learning, Instruction & Teacher Education, University of Haifa, Israel // Ina.Blau@openu.ac.il // azy@edu.haifa.ac.il

ABSTRACT

The development of digital technologies increases the use of distance synchronous (real-time) interactions among people. The study explores whether the readiness to participate, the degree of actual participation, and the quality of contribution to synchronous online group discussions is affected by participant personality, media characteristics, and discussion topic sensitivity. The relation between anticipated and actual participation was investigated, as well as the interpersonal and gender equalization effects of online communication. An online self-report viral survey was completed by 405 adult Internet users. Following that, 120 volunteers extracted from this sample were randomly assigned to small, gender-mixed groups, employing face-to-face, online audio or online text chat experimental conditions, and conducted two non-moderated discussions (having low and high degrees of topic sensitivity). A greater interest in discussing sensitive over non-sensitive topic can explain higher participation and quality of contribution. Online text chat appeared as an efficient medium, in which the quality of participant contributions was similar to spoken discussions, obtained by smaller amount of words. Discussing sensitive topic, participants felt more comfortable using "lean" text-only medium. As hypothesized, participant personality affected the involvement in discussions: extroverts preferred taking part via a more revealing communication medium while introverts expressed greater readiness for holding discussions via text chat.

Keywords

Synchronous participation, Extroversion-introversion personality trait, Discussion topic, Audio and text chat, Interpersonal and gender equalization effect

Introduction

In recent years, online communication technologies have penetrated almost every aspect of our life and have become an essential element in interpersonal interactions, including study and work environments. The fast development of digital communication increases technology-mediated participation and the use of online synchronous communication among people. Synchronous technologies enable real-time interpersonal interactions in cyberspace that creates a social environment based on Internet infrastructure (Barak & Suler, 2008). These online interpersonal interactions are held either through one-to-one (personal) or group communication. This study explores some of the psychological aspects of synchronous group interactions through the Internet. We argue that involvement in online group discussions is affected, among other factors, by participant personality, medium characteristics, and discussion topic. Our study investigates the degree of quantitative and qualitative participation in face-to-face (FiF) communication versus technology-mediated interactions through audio chat and text chat held in small discussion groups. The study explores whether the readiness to participate, as well as the degree of actual involvement in group discussions, is affected by participant personality (i.e., extroversion-introversion), medium characteristics, and the sensitivity of discussion topic. In addition, we investigated whether online participation is more balanced compared to offline one, because of the diminished cues for social status and the gender differences between communicators.

Related Studies

There are several theories that differentiate media by their inherent features in order to predict an efficient communication for interpersonal and group interactions. One of the leading theoretical approaches is Media Richness Theory (MRT; Daft & Lengel, 1984). According to MRT, communication media varies in their ability to transmit social communication cues; FiF communication—the "richest" medium—is considered the most efficient way to convey complex messages. Empirical research regarding the influence of media richness on communication provides mixed results: some studies found evidence supporting it (see review by Donabedian, 2006); other findings pointed to the fact that distractors provided by rich medium may draw attention away from the transmitted message and
negatively affect communication (Blau & Caspi, 2008, 2010; Setlock, Quinones & Fussell, 2007). Some researchers claimed that a "lean" communication medium (e.g., e-mail or text chat) may still provide rich interpersonal interactions. Walther's (1996, 2007) findings on the "hyperpersonal effect" of e-communication showed that text-based interactions can possibly equal or even exceed relational effects derived from comparable FtF interactions. This approach received extensive theoretical and empirical support in studying online communication (e.g., Joinson, 2007; Suler, 2004) and e-learning (e.g., O'Sullivan, Hunt, & Lippert, 2004). However, Walther's claim that online video, audio, text, or future applications may be as efficient as FtF communication (Walther, Loh & Granka, 2005) was empirically studied mostly in text environments. The increasing use of online audio and video communication (Ng, 2007) requires investigating the impact of these media on online behavior of participants. As a normal and a primary communication ingredient, voice provides meta-communication features of human messages through loudness, intonation, pitch, and breaks (Pickett, 1998). According to the Media Naturalness Theory, the ability to transmit human voice is critical for determining the degree of medium naturalness (Kock, 2009; Kock, Chatelain-Jardón, & Carmona, 2008). Therefore, voice chat may have a different effect on interpersonal interactions than text chat (Barak, 2007) because of the different degrees of anonymity (visual anonymity versus both visual and auditory anonymity), and may result in different group behavior (Blau & Caspi, 2007).

Studies comparing participation through FtF and synchronous online communication revealed inconsistent results: while some of them found decreased participation in voice and text chat (Berge & Fjuk, 2006), others claimed that text chat (Hudson & Bruckman, 2002; Lobel et al., 2002) and audio communication (Blau & Caspi, 2008, 2010) increase participation compared to FtF interactions.

Personality characteristics of users may impact interpersonal and group communication; the trait of extroversion-introversion is especially relevant to online behavior (Amichai-Hamburger, 2007, Amichai-Hamburger & Barak, 2009). An extrovert is a friendly person who seeks company, desires excitement, takes risks, and acts on impulse, whereas an introvert is a quiet, reflective person who does not enjoy large social events, prefers his or her own company, and does not crave excitement (Eysenck & Eysenck, 1975). Discussing the Internet as a compensatory psychological tool, Kraut et al. (2002) claimed that people who easily befriend offline would also have more online friends (so called "the rich get richer" phenomenon). Other researchers hold that the protected Internet environment may assist introverts in expressing themselves more freely in an online than in offline relationships ("the poor get richer"; Amichai-Hamburger, 2007; Maldonado et al., 2001; McKenna, Green & Gleason, 2002). While extroverts feel comfortable both online and offline, introverts express themselves significantly more freely over the Internet (McKenna, Seidman, Buffardi & Green, 2007). Interacting online, introverts even adopt offline behavior patterns of extroverts (Amichai-Hamburger, Wainapel & Fox, 2002; Maldonado et al., 2001; McKenna & Seidman, 2005) which could be reflected in more active participation in online rather than in offline discussions (McKenna et al., 2007).

Research shows that the relationships between people's personality traits and their online behavior may be moderated by type and amount of the Internet use (Anolli, Villani & Riva, 2005), participant demographics (Amichai-Hamburger & Ben-Artzi, 2000; Maldonado et al., 2001), and the sensitivity of discussion topic (Hertel, Schroer, Batinic & Naumann, 2008). However, the influence of topic sensitivity on the readiness to partake in online discussions was examined and studied by these researchers in asynchronous communication environment and may differ from real-time interactions.

In most cases, compared to offline interactions, online environment provides more equal opportunities to people to voice themselves, regardless of status, gender, race, wealth, or appearance (Suler, 2004). Online group interactions diminish external and internal, real and fictitious status cues (Amichai-Hamburger, 2007; Amichai-Hamburger & Barak, 2009; Barak, Boniel-Nissim, & Suler, 2008; McKenna, 2008). This equalization effect of online communication was found in several laboratory experiments (Dubrovsky, Kiesler, & Sethna, 1991; Siegel, Dubrovsy, Kiesler, & McGuire, 1986) and field studies (Warschauer, 1996). These studies, however, tested the equalization effect through textual communication, which may differ from equalization in spoken online interactions.

Study Goals and Hypotheses

This study explores some psychological aspects of synchronous group interactions. Group behavior in FtF communication versus online interactions through audio chat and text chat were compared. Specifically, the study
examined whether the readiness to participate, the degree of actual involvement, and the quality of contribution to discussions are affected by participants' personality (i.e., extroversion-introversion), communication mode (i.e., F2F, online voice chat, or online text chat), and the sensitivity of discussion topic (low versus high). The readiness to partake in discussions reported by participants was crosschecked with the degree of their actual participation. In addition, the interpersonal and gender equalization effect of online communication was explored.

We hypothesized that extroversion-introversion would affect the degree of anticipated and actual participation as well as the quality of contribution to discussions in F2F communication versus online voice chat or text chat interactions, and that the sensitivity of discussion topic would augment this trend. In addition, we hypothesized that the readiness to participate would positively correlate with the degree of actual involvement in discussions. Concerning the equalization effect, we hypothesized that, in terms of both interpersonal and gender equality, online participation would be more equal compared to offline ones, and the participation using text chat would be more equal in comparison to participation through audio chat.

The research is composed of two sequenced sub-studies.

Study 1

The first study investigates the readiness to participate in discussion as related to medium characteristics, participant personality, and the sensitivity of discussion topic.

Method

Participants

An online viral sampling of 405 Israeli adult Internet users, 118 of them men (29.1%), was composed. Participants' age range was 18-76, median 29, mean 31.89, SD 10.97. Level of education: 37.5% were high-school graduates, 9.9% had professional post-school training, 31.9% held BA degrees, and 20.7% held masters or doctoral degrees. Similar (relatively high) percentage of graduate participants (38%) was reported by Mesch and Elgali (2009) in a survey of 1,000 Israeli Internet users. Table 1 presents the participant self-reported usage of the Internet in general and synchronous communication tools in particular.

<table>
<thead>
<tr>
<th>Usage level</th>
<th>General Internet usage</th>
<th>Personal chat</th>
<th>Chat room</th>
<th>Audio communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each day</td>
<td>85.7</td>
<td>24.4</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>2-3 times a week</td>
<td>10.6</td>
<td>14.6</td>
<td>0.7</td>
<td>6.2</td>
</tr>
<tr>
<td>Once a week- Once a month</td>
<td>3.7</td>
<td>13.8</td>
<td>3.2</td>
<td>9.4</td>
</tr>
<tr>
<td>Less than once a month</td>
<td>0</td>
<td>15.1</td>
<td>6.9</td>
<td>18</td>
</tr>
<tr>
<td>Not using</td>
<td>0</td>
<td>31.9</td>
<td>87.2</td>
<td>62.7</td>
</tr>
</tbody>
</table>

Instruments

The participants completed an online self-report questionnaire, which include two parts. In order to measure the trait of extraversion-introversion we used a Hebrew short version of the NEO-PI-R questionnaire based on the Big Five model (Costa & McCrae, 1992). The score of extroversion-introversion (Range=21-69, Medium=51, Mean=51.26, SD=8.08) was computed by summing up participant answers to 12 items (scale 1-6); internal consistency α=.80. Participants were divided into two groups (extroverts vs. introverts) using the median scale score (51; the participants who received the median score were defined as introverts). In addition, following Hertel et al. (2008), the participants were asked to imagine first group discussion on non-sensitive topic ("Please imagine chatting with a friend, colleague, or peer about the last weekend"), then the sensitive one ("Please imagine you are angry about the egoistic behavior of a friend, colleague, or peer"), and report about their readiness to discuss F2F, through audio and text chat (scale 1 = "not at all"; 8 = "very much").
Procedure

Online viral sampling method was used: a link to online self-report questionnaire was sent to participants by email asking for resending the link to their contacts and partaking in further experiments.

Results

Effects of Media and Discussion Topic on the Readiness to Participate

In order to test the study hypotheses, series of ANOVA Repeated Measures tests were conducted. Table 2 shows the means and standard deviations of the participants' readiness to discuss the non-sensitive and sensitive topics through FtF, audio and text chat.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Medium→</th>
<th>FtF</th>
<th>Audio</th>
<th>Chat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-sensitive</td>
<td>M</td>
<td>6.60</td>
<td>3.15</td>
<td>3.23</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.57</td>
<td>1.90</td>
<td>1.95</td>
<td>1.06</td>
</tr>
<tr>
<td>Sensitive</td>
<td>M</td>
<td>6.06</td>
<td>2.66</td>
<td>3.06</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.98</td>
<td>1.77</td>
<td>2.26</td>
<td>1.03</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>6.33</td>
<td>2.90</td>
<td>3.14</td>
<td>4.13</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.43</td>
<td>1.63</td>
<td>1.80</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 3 presents results of the ANOVA for the effects of communication medium, discussion topic, and the interaction between the two on the readiness to participate in discussions.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Partial ( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>483.1</td>
<td>2, 403</td>
<td>&lt;.001</td>
<td>.71</td>
</tr>
<tr>
<td>Topic</td>
<td>62.02</td>
<td>1, 404</td>
<td>&lt;.001</td>
<td>.13</td>
</tr>
<tr>
<td>Medium x Topic</td>
<td>3.45</td>
<td>2, 403</td>
<td>&lt;.05</td>
<td>.03</td>
</tr>
</tbody>
</table>

Significant main effects were found for communication medium, the sensitivity of the discussion topic, and their interaction on participant readiness to partake in discussions. Generally, participants preferred the non-sensitive over the sensitive discussion topic (M=4.33 vs. 3.93, \( p < .001 \)). Post-hoc tests showed that they also preferred FtF discussion rather than audio chat or text chat (M=6.33 vs. 2.90 and 3.14, \( p < .001 \)). For online interactions, participants preferred using text chat rather than audio communication (\( p < .05 \)). In regard to discussing a non-sensitive topic, there were no significant differences between audio and text chat, but offline interaction was preferable than both online media (M=3.15 and 3.23 vs. 6.60, \( p < .001 \)). Concerning the online discussions of a sensitive topic, text chat was preferable over the audio chat (M=3.06 vs. 2.66, \( p < .001 \)). Is seems that participants felt more comfortable discussing the sensitive topic via a "lean" communication medium.

Effect of Extroversion-Introversion on the Readiness to Participate

Table 4 shows the means and standard deviations of extrovert and introvert readiness to discuss non-sensitive and sensitive topics using different communication media.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Medium→</th>
<th>F2F</th>
<th>Audio</th>
<th>Chat</th>
<th>Total</th>
<th>F2F</th>
<th>Audio</th>
<th>Chat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-sensitive</td>
<td>M</td>
<td>6.88</td>
<td>3.41</td>
<td>3.21</td>
<td>4.50</td>
<td>6.34</td>
<td>2.91</td>
<td>3.25</td>
<td>4.17</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.36</td>
<td>1.90</td>
<td>1.87</td>
<td>1.05</td>
<td>1.70</td>
<td>1.87</td>
<td>2.03</td>
<td>1.06</td>
</tr>
<tr>
<td>Sensitive</td>
<td>M</td>
<td>6.39</td>
<td>2.83</td>
<td>3.01</td>
<td>4.08</td>
<td>5.75</td>
<td>2.50</td>
<td>3.11</td>
<td>3.79</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.86</td>
<td>1.77</td>
<td>2.21</td>
<td>0.96</td>
<td>2.04</td>
<td>1.75</td>
<td>2.30</td>
<td>1.06</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>6.33</td>
<td>3.12</td>
<td>3.11</td>
<td>4.29</td>
<td>6.05</td>
<td>2.70</td>
<td>3.18</td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.43</td>
<td>1.60</td>
<td>1.72</td>
<td>0.86</td>
<td>1.54</td>
<td>1.63</td>
<td>1.87</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Table 5 presents the ANOVA for the effects of extroversion-introversion, type of communication medium, the sensitivity of discussion topic, and their interactions, on the readiness to participate in discussions.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extroversion</td>
<td>12.14</td>
<td>1, 404</td>
<td>&lt;.001</td>
<td>.05</td>
</tr>
<tr>
<td>Type of Medium</td>
<td>484.26</td>
<td>2, 403</td>
<td>&lt;.001</td>
<td>.71</td>
</tr>
<tr>
<td>Topic Sensitivity</td>
<td>62.01</td>
<td>1, 404</td>
<td>&lt;.001</td>
<td>.13</td>
</tr>
<tr>
<td>Extroversion x Medium</td>
<td>3.81</td>
<td>2, 403</td>
<td>&lt;.05</td>
<td>.03</td>
</tr>
<tr>
<td>Extroversion x Topic</td>
<td>0.14</td>
<td>1, 404</td>
<td>n.s.</td>
<td>.00</td>
</tr>
<tr>
<td>Medium x Topic</td>
<td>3.46</td>
<td>2, 403</td>
<td>&lt;.05</td>
<td>.03</td>
</tr>
<tr>
<td>Extroversion x Medium x Topic</td>
<td>0.54</td>
<td>2, 403</td>
<td>n.s.</td>
<td>.00</td>
</tr>
</tbody>
</table>

Participant level of extroversion affected the readiness to partake in discussions: extroverts were more inclined to participate in discussions than introverts ($M=4.29$ vs. $3.98$, $p<.001$). The interaction between participant extroversion and type of medium revealed that extroverts were more inclined to partake compared to introverts F2F ($M=6.63$ vs. $6.05$, $p<.001$) and through audio chat ($M=3.12$ vs. $2.70$, $p<.01$), but not through text chat. For online discussions, extroverts did not show preferences for a particular communication medium, while introverts preferred text chat over audio chat ($M=3.18$ vs. $2.70$, $p<.01$). Thus, introverts expressed greater readiness than extroverts for discussions via medium low in transmitting social communication cues.

### Study 2

In this study we explored the degree of actual participation in a group discussion and evaluated the quality of contribution to discussions, as influenced by participant level of extraversion-introversion, communication medium, and sensitivity level of the discussion topic. In addition, we examined the interpersonal and gender equalization effect of online communication, as well as the relationship between anticipated and actual participation in discussions.

### Method

#### Participants

The participants of the second study consisted of 120 volunteers derived from the sample generated in Study 1. Forty eight (40%) of the participants were men.

#### Manipulations

We manipulated the level of sensitivity of the discussion topics by offering two topics for discussions. For low-level sensitivity discussions we adapted the topic from Warschauer's (1996) study: "If a man and a woman are living together and each works a full-time job (40 hours a week), how much and what type of housework should each of them do?" For high-level sensitivity discussions the topic was: "Think about relationships about couples; which online or offline behavior would you characterize as unfaithfulness and for what reason?"

#### Measures

To enhance external validity, measurement of three of the dependent variables was conducted by using ethnographic, non-obtrusive and nonreactive, behavioral observations rather than through self-report questionnaires (Fritsche & Linneweber, 2006). For amount of actual participation in discussions we counted the number of words used by each participant (Range=18-1029, Median=150, Mean=184.72, SD=100). Interpersonal equalization was measured by participation percentage (i.e., number of words per participant divided by number of words expressed (said or written) by all group members during the discussion; Blau & Caspi, 2007; Warschauer, 1996), Range=1.27-47.84,
Median=18.86, Mean=20.00, SD=10.26. Gender equalization was measured by gender participation percentage (i.e., number of words expressed (said or written) by male/female participant divided by number of words of all group members during the discussion; Caspi, Chajut, & Saporta, 2008). The quality of contribution to discussions for each participant was assessed by average evaluation of four expert judges "blind" to the hypotheses (on a scale of 1 = "not at all"; 5 = "very much", Median=3.00, Mean=3.01, SD=1.00); between-rater agreement coefficient Kendall's W=.70. The evaluation of contribution was based on suggesting new ideas, definitions, focusing, and summarizing discussions.

Procedure

Participants were randomly assigned to gender-mixed experimental conditions and conducted, in groups of five, two short (average of 18.40 min.) non-moderated discussions on both low and high degree of topic sensitivity, 40 participants in each communication mode. Similarly to online natural discussions, each discussion in this study was non-moderated and was terminated when the participants felt they fully expressed their opinion. The study did not use the counterbalance procedure; instead, all groups discussed non-sensitive topic before the sensitive one. Discussing the sensitive topic first could have left the participants exited and neutralize the difference between the topics. Skype™ application was used for online discussions, either for text or audio chat. The discussions were recorded, transcribed, and evaluated by four expert judges for the quality of contribution for each participant. The amount of actual participation, as well as interpersonal and gender equality were calculated. The readiness to partake in discussions was compared to the actual behavior of the participants.

Results

Effects of Media and Discussion Topic on Actual Participation

Table 6 shows the means and standard deviations for the amount of actual participation (number of words per participant) in discussing non-sensitive and sensitive topics FtF, through audio and text chat.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Medium→</th>
<th>F2F (n=40)</th>
<th>Audio (n=40)</th>
<th>Chat (n=40)</th>
<th>Total (n=120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-sensitive</td>
<td>M</td>
<td>150.63</td>
<td>154.90</td>
<td>109.45</td>
<td>138.33</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>124.71</td>
<td>120.75</td>
<td>86.98</td>
<td>116.74</td>
</tr>
<tr>
<td>Sensitive</td>
<td>M</td>
<td>281.73</td>
<td>273.55</td>
<td>138.05</td>
<td>231.11</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>169.56</td>
<td>120.83</td>
<td>105.52</td>
<td>121.38</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>216.18</td>
<td>214.23</td>
<td>123.75</td>
<td>184.72</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>116.82</td>
<td>108.92</td>
<td>87.47</td>
<td>100.40</td>
</tr>
</tbody>
</table>

Table 7 presents the analysis of variance for the effects of communication medium, discussion topic, and the interaction between the two on the amount of actual participation in discussions.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Partial η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>5.29</td>
<td>2, 117</td>
<td>&lt;.01</td>
<td>.08</td>
</tr>
<tr>
<td>Topic</td>
<td>24.33</td>
<td>1, 117</td>
<td>&lt;.001</td>
<td>.17</td>
</tr>
<tr>
<td>Medium x Topic</td>
<td>3.05</td>
<td>2, 117</td>
<td>&lt;.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

Significant main effects were found for communication medium, the sensitivity of discussion topic, and their interaction. Generally, actual involvement was greater for the sensitive over non-sensitive discussion topic (M=231.11 vs. 138.33, p<.01). Post-hoc tests showed no difference between FtF and audio chat discussions (M=216.82 and 214.23, respectively); however, participation in audio chat was significantly higher compared to text chat (M=214.23 vs. 123.75, p<.01). The interaction effect showed more active participation in sensitive compared to non-sensitive topic for FtF (M=281.73 vs. 150.63, p<.01) and audio discussions (M=273.55 vs. 154.90, p<.01), but not for text chat (M=138.05 vs. 109.45, p>.05). Discussing a sensitive topic, participation was significantly smaller in text chat compared to FtF and to audio chat (M=138.05 vs. 281.73 and vs. 273.55, respectively, both p<.01).
However, in discussing a non-sensitive topic there was no significant differences among types of media that were found.

**Effect of Extroversion-Introversion on Actual Participation**

Table 8 shows the means and standard deviations of the actual participation discussing non-sensitive and sensitive topics by different communication media by level of extroversion.

<table>
<thead>
<tr>
<th></th>
<th><strong>F2F (n=23)</strong></th>
<th><strong>Audio (n=22)</strong></th>
<th><strong>Chat (n=16)</strong></th>
<th><strong>Total (n=61)</strong></th>
<th><strong>F2F (n=17)</strong></th>
<th><strong>Audio (n=18)</strong></th>
<th><strong>Chat (n=24)</strong></th>
<th><strong>Total (n=59)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extroverts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-sensitive</td>
<td>198.00</td>
<td>167.55</td>
<td>123.93</td>
<td>167.59</td>
<td>86.53</td>
<td>139.44</td>
<td>99.79</td>
<td>108.07</td>
</tr>
<tr>
<td>Sensitive</td>
<td>300.00</td>
<td>337.68</td>
<td>164.56</td>
<td>278.07</td>
<td>257.00</td>
<td>195.17</td>
<td>120.38</td>
<td>182.56</td>
</tr>
<tr>
<td><strong>Introverts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-sensitive</td>
<td>98.31</td>
<td>98.39</td>
<td>70.17</td>
<td>81.12</td>
<td>46.70</td>
<td>95.04</td>
<td>66.81</td>
<td>64.66</td>
</tr>
<tr>
<td>Sensitive</td>
<td>108.12</td>
<td>156.52</td>
<td>82.97</td>
<td>109.89</td>
<td>104.42</td>
<td>99.41</td>
<td>85.60</td>
<td>82.56</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>249.00</strong></td>
<td><strong>252.61</strong></td>
<td><strong>144.25</strong></td>
<td><strong>222.83</strong></td>
<td><strong>171.76</strong></td>
<td><strong>167.31</strong></td>
<td><strong>110.08</strong></td>
<td><strong>145.30</strong></td>
</tr>
</tbody>
</table>

Table 9 presents the ANOVA for the effects of extroversion-introversion, communication mode, the sensitivity of discussion topic, and their interactions on the amount of actual participation in discussions.

<table>
<thead>
<tr>
<th><strong>Effect</strong></th>
<th><strong>F</strong></th>
<th><strong>df</strong></th>
<th><strong>p</strong></th>
<th><strong>Partial (\eta^2)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Extroversion</td>
<td>6.17</td>
<td>1, 114</td>
<td>&lt;.05</td>
<td>.06</td>
</tr>
<tr>
<td>Type of Medium</td>
<td>4.36</td>
<td>2, 114</td>
<td>&lt;.05</td>
<td>.07</td>
</tr>
<tr>
<td>Topic Sensitivity</td>
<td>24.24</td>
<td>1, 114</td>
<td>&lt;.001</td>
<td>.18</td>
</tr>
<tr>
<td>Extroversion x Medium</td>
<td>0.36</td>
<td>2, 114</td>
<td>n.s.</td>
<td>.01</td>
</tr>
<tr>
<td>Extroversion x Topic</td>
<td>0.34</td>
<td>1, 114</td>
<td>n.s.</td>
<td>.00</td>
</tr>
<tr>
<td>Medium x Topic</td>
<td>3.04</td>
<td>2, 114</td>
<td>&lt;.05</td>
<td>.05</td>
</tr>
<tr>
<td>Extroversion x Medium x Topic</td>
<td>3.06</td>
<td>4, 114</td>
<td>&lt;.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

Participant extroversion level affected the degree of involvement in discussions: extroverts engaged in more active participation in discussions than introverts (M=222.83 vs. 145.30, \(p<.05\)). The interaction effect between level of extroversion and communication mode, and between level of extroversion and the sensitivity of the discussion topic on amount of participation did not reveal significant results. The triple interaction showed that extroverts were more active through audio chat discussing sensitive rather than non-sensitive topic (M=337.68 vs. 167.55, \(p<.01\)), while no similar difference was found among introverts. Discussing the non-sensitive topic, extroverts showed greater activity through F2F compared to text chat (M=198.00 vs. 123.93, \(p<.05\)); however, no significant difference was found among introverts.

**Effects of Media and Discussion Topic on Quality of Contribution to Discussion**

Table 10 shows the means and standard deviations for mean rater evaluations for the quality of contribution to discussions on non-sensitive and sensitive topics through F2F, audio and text chat.

<table>
<thead>
<tr>
<th><strong>Topic</strong></th>
<th><strong>Medium →</strong></th>
<th><strong>F2F (n=40)</strong></th>
<th><strong>Audio Chat (n=40)</strong></th>
<th><strong>Text Chat (n=40)</strong></th>
<th><strong>Total (n=120)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-sensitive</td>
<td>M</td>
<td>2.54</td>
<td>2.91</td>
<td>2.70</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.02</td>
<td>1.30</td>
<td>1.10</td>
<td>1.14</td>
</tr>
<tr>
<td>Sensitive</td>
<td>M</td>
<td>3.29</td>
<td>3.41</td>
<td>3.18</td>
<td>3.20</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.24</td>
<td>1.39</td>
<td>1.17</td>
<td>1.26</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>M</td>
<td><strong>2.91</strong></td>
<td><strong>3.16</strong></td>
<td><strong>2.94</strong></td>
<td><strong>3.00</strong></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td><strong>0.92</strong></td>
<td><strong>1.21</strong></td>
<td><strong>0.84</strong></td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>
Table 11 presents the ANOVA for the effects of communication mode, discussion topic, and the interaction between the two on the quality of participant contribution to discussions.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>0.75</td>
<td>2, 117</td>
<td>n.s.</td>
<td>.01</td>
</tr>
<tr>
<td>Topic</td>
<td>22.21</td>
<td>1, 117</td>
<td>&lt;.001</td>
<td>.16</td>
</tr>
<tr>
<td>Medium x Topic</td>
<td>0.52</td>
<td>2, 117</td>
<td>n.s.</td>
<td>.01</td>
</tr>
</tbody>
</table>

Significant main effect was found for the level of sensitivity of the discussion topic: higher quality of participant contribution to discussions was rated for the sensitive compared to non-sensitive discussion topic (M=3.29 vs. 2.72, $p<.001$). No significant effects were found neither for communication medium nor the medium-topic interaction.

Effect of Extroversion-Introversion on Quality of Contribution to Discussion

Table 12 shows the means and standard deviations for the quality of contribution discussing non-sensitive and sensitive topics by communication mode and extroversion-introversion.

Table 12. Means and SDs of Quality of Contribution by Extroversion-Introversion, Medium, and Topic

<table>
<thead>
<tr>
<th>Medium → Topic ↓</th>
<th>Extroverts</th>
<th>Introverts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F2F (n=23)</td>
<td>Audio (n=22)</td>
</tr>
<tr>
<td>Non-sensitive M</td>
<td>2.93</td>
<td>3.09</td>
</tr>
<tr>
<td>SD</td>
<td>0.83</td>
<td>1.20</td>
</tr>
<tr>
<td>Sensitive M</td>
<td>3.35</td>
<td>3.68</td>
</tr>
<tr>
<td>SD</td>
<td>1.31</td>
<td>1.26</td>
</tr>
<tr>
<td>Total M</td>
<td>3.14</td>
<td>3.39</td>
</tr>
<tr>
<td>SD</td>
<td>0.87</td>
<td>1.12</td>
</tr>
</tbody>
</table>

Table 13 presents the ANOVA for the effects of extroversion-introversion, communication mode, the degree the sensitivity of discussion topic, and their interactions, on the quality of participant contribution in discussions.

Table 13. Effects of Extroversion-Introversion, Medium, and Topic on the Quality of Contribution

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extroversion</td>
<td>8.41</td>
<td>1, 114</td>
<td>&lt;.01</td>
<td>.07</td>
</tr>
<tr>
<td>Type of Medium</td>
<td>0.72</td>
<td>2, 114</td>
<td>n.s.</td>
<td>.01</td>
</tr>
<tr>
<td>Topic Sensitivity</td>
<td>22.34</td>
<td>1, 114</td>
<td>&lt;.001</td>
<td>.16</td>
</tr>
<tr>
<td>Extroversion x Medium</td>
<td>0.91</td>
<td>2, 114</td>
<td>n.s.</td>
<td>.01</td>
</tr>
<tr>
<td>Extroversion x Topic</td>
<td>1.54</td>
<td>1, 114</td>
<td>n.s.</td>
<td>.02</td>
</tr>
<tr>
<td>Medium x Topic</td>
<td>0.88</td>
<td>2, 114</td>
<td>n.s.</td>
<td>.01</td>
</tr>
<tr>
<td>Extroversion x Medium x Topic</td>
<td>3.08</td>
<td>4, 114</td>
<td>&lt;.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

Extroverts significantly affected the quality of participant contributions in the discussions: contribution of extroverts was greater compared to introverts (M=3.26 vs. 2.74, $p<.01$). The interaction effect between extroversion and communication mode, as well as between extroversion and sensitivity of the discussion topic, on quality of participant contribution to discussions was not found to be significant. The triple interaction showed that for introverts the quality of contribution to discussions was greater in FtF discussing sensitive rather than non-sensitive topic (M=3.21 vs. 2.00, $p<.01$), while no such a difference was found among extroverts. Discussing non-sensitive topic, the quality of contribution for introverts was greater through audio chat compared to FtF mode (M=2.69 vs. 2.00, $p<.05$); however, no similar difference was found among extroverts.

Interpersonal Equalization of Participation

Since interpersonal equalization is measured by participation percentage for each participant in group of five, average participation percentage is always 20% and therefore meaningless. The comparison between standard
deviations showed smaller variance, i.e., more equal participation, for text chat compared to F2F and discussions through audio chat (SD=8.20 vs. 12.77 and 12.89 respectively, n=40 for each medium). Levene's test for equality of variances showed significant differences between the equality of participation through text chat compared to F2F, as well as between text chat compared to audio chat (Levene's W=6.94 and 7.46 respectively, p's<.01).

Gender Equalization of Participation

Table 14 presents the means and standard deviations for the percentage of participation in discussions in the different communication modes (regardless of the discussion topic) by gender.

Table 14. Means and SDs of Participation in Discussions by Gender and Medium

<table>
<thead>
<tr>
<th>Gender ↓</th>
<th>Medium→</th>
<th>FtF (n=40)</th>
<th>Audio Chat (n=40)</th>
<th>Text Chat (n=40)</th>
<th>Total (n=120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (n=48)</td>
<td>M</td>
<td>22.84</td>
<td>27.39</td>
<td>16.89</td>
<td>23.81</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>12.61</td>
<td>12.76</td>
<td>8.04</td>
<td>10.61</td>
</tr>
<tr>
<td>Female (n=72)</td>
<td>M</td>
<td>19.29</td>
<td>17.86</td>
<td>21.66</td>
<td>19.30</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>12.06</td>
<td>11.79</td>
<td>8.17</td>
<td>10.06</td>
</tr>
</tbody>
</table>

Table 15 presents results of the ANOVA for the effects of gender, communication mode, and the interaction between the two on the percentage of participation to discussions.

Table 15. Effects of Gender and Medium on the Percentage of Participation in Discussions

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Partial η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>3.08</td>
<td>5, 234</td>
<td>&lt;.05</td>
<td>.03</td>
</tr>
<tr>
<td>Medium</td>
<td>1.48</td>
<td>5, 234</td>
<td>n.s</td>
<td>.01</td>
</tr>
<tr>
<td>Gender x Medium</td>
<td>4.40</td>
<td>5, 234</td>
<td>&lt;.01</td>
<td>.04</td>
</tr>
</tbody>
</table>

Significant main effect was found for gender: men participated more actively compared to women (M=23.81 vs. 19.30, p<.05). Examining the significant interaction effect it was found that there was no gender difference in the FtF mode; however, men participated more than women in audio chat (M=27.39 vs. 17.86, p<.01), in contrast women participated more actively through text chat (M=21.66 vs. 16.89, p<.05).

Readiness to Participate versus Actual Participation

The readiness to partake in discussions as reported by participants in Study 1 was crosschecked with the degree of their actual participation in Study 2. Surprisingly, Pearson correlations for the readiness to participate in FtF, audio chat, and text chat with the degree of actual participation in discussions did not reveal statistical significance (r=.13, 14, and .02 respectively, n=40 for each medium).

General Discussion

Online communication technologies are altering the traditional modes of interpersonal communication for learning and working, information consumption and creation, discussions and conversations, and knowledge construction. This study investigated the level of anticipated and actual participation, as well as the quality of personal contribution in group discussions. Discussing non-sensitive versus sensitive topic, FtF communication mode was compared to synchronous interactions through audio chat and text chat. In addition, the study explored the equalization effect of online communication compared to offline one.

Generally, the amount of actual participation and the quality of contribution to discussions were found to be greater for the sensitive over non-sensitive discussion topic. Apparently, this finding can be explained by participants' greater interest in the discussions of a sensitive topic. Based on this explanation, the use of sensitive, intriguing, and challenging topics is recommended for increasing involvement in group discussions in various social environments, including learning and work. It should be kept in mind, however, that since we could not use the counterbalance procedure in our research, there is still a rival possible explanation that a high amount of participation (and quality of
contribution) was created by the order of the discussion conditions. We recommend continued testing of the amount of actual involvement in sensitive versus non-sensitive topics in future studies.

While expressing the readiness to discuss a sensitive topic, the participants preferred a text chat over a voice chat. It seems that participants felt more comfortable discussing sensitive topics via a "lean" communication medium. The analysis of actual behavior showed that the amount of participation through text chat was smaller compared to other communication modes; this may reinforce emotional processing and reflection on the topic. Thus, for discussing sensitive topics in real-time in various interpersonal environments, employment of lean communication modes, such as text chat, is recommended.

However, across discussion subjects, smaller amount of actual participation was revealed through text chat in comparison with spoken F2F and audio chat discussions. According to Media Naturalness Theory, the ability to transmit human voice is critical for determining the degree of medium naturalness (Kock, 2009; Kock et al., 2008). Thus, voice chat has a different effect on interpersonal interactions than text chat because of the different degrees of anonymity (visual anonymity versus both visual and auditory anonymity), therefore resulting in higher levels of participation. Thus, audio chat seems to be a more appropriate medium than text chat for encouraging involvement in synchronous communication if the subject of discussion cannot be manipulated.

Lower degree of participation can be also explained by the effect of the well-known phenomenon of information overload (Hiltz & Turoff, 1985), shown to occur in text chat (Jones, Moldovan, Raban, & Butler, 2008). According to these researchers, information overload takes place because of transmitting too many messages and/or the lack of structure and coherence, which leads to diminished participant activity. Despite the advantages of chat explored in this study, textual group discussions are less structured compared to spoken ones. In contrast to clear turn-taking for speaking in F2F and audio chat communication modes, in text chats participants express themselves (i.e., write messages) simultaneously. Chatters therefore needed higher concentration in order to express their opinions and to follow parallel arguments by others, which frequently lead to diminished amounts of participation. However, the quality of contribution found in text chat discussions in the current study was similar to the spoken F2F and audio chat communication modes. Thus, text chat can be seen as an efficient form of real-time communication, in which the quality of the contribution was similar to spoken discussions obtained by shorter messages (i.e., fewer words).

As hypothesized, personality affected the readiness to participate in discussions: extroverts expressed greater readiness to participate in discussions compared to introverts. Also, this personality trait interacted with the type of medium in affecting the readiness to participate in discussions: introverts expressed greater readiness to participate in discussions via text chat compared to extroverts (who preferred more revealing communication media). Similar results revealed in analyzing the influence of extroversion on participants' actual behavior. As expected, extroverts participated in discussions significantly more actively than introverts. This finding is consistent with previous studies that reported extroverts' higher involvement in interpersonal interactions. In the current study, participation of introverts in F2F communication was very limited; a sensitive discussion topic, however, augmented the rate of participation. Interestingly, the interaction effect found between extroversion and type of media revealed that the level of participation of extroverts and introverts was similar in text chat but differed in the more natural communication modes. It seems that in synchronous communication, only text chat (i.e., low in communication richness) empowers introverts by releasing inhibitions of their anxiety of interpersonal interactions ("the poor get richer" hypothesis). However, as complicated interactions of introversion-extroversion with online communication may exist (Valkenburg & Peter, 2007), this subject should be examined more closely in future research.

Participant extroversion affected not only the amount of participation but also the quality of the discussions. The quality of extroverts’ contribution to discussions, as evaluated by raters, was higher compared to the contribution of introverts. This finding has important implications for the evaluation of real-time group discussions in learning and work environments. That is, introvert learners or employees may have good ideas, but feel uncomfortable expressing them in group meetings. In this case follow-up asynchronous reflections through email, blog, or forum formats are recommended.

Our results showed gaps between the participant anticipated and actual behavior. In contrast to our hypothesis, there were no significant correlations between the readiness to partake in discussions as reported by participants in Study 1 and the amount of their actual participation in discussions as assessed in Study 2. Also, despite the clear preference for non-sensitive over sensitive discussion topics declared by participants before the experiment (Study 1), their
actual participation discussing sensitive topics was almost twice as high as discussing non-sensitive topic (Study 2). In addition, participants declared they would prefer taking part in discussions via FtF over online communication modes (Study 1); however, almost the same amount of participation was found in actual FtF and audio chat discussions (Study 2). Similarly, in expressing preferences for online communication, participants reported greater readiness to interact through text chat compared to audio communication (Study 1). However, their actual spoken participation in audio chat was almost twice as much as text chat participation (Study 2). These findings question the possibility of studying actual participant behavior by self-reported anticipations of future selections and decisions.

Regarding the equalization effect, consistent with our hypothesis, more equal interpersonal and gender-related participation was found in text chat compared to FtF and to audio communication modes. However, equality of participation in voice chat was similar to equality of participation in FtF discussions. It seems that human speech transmits important status- and gender-related cues and therefore disables the occurrence of equalization effect in spoken online interactions. Thus, in real-time communication the use of text chat might be recommended in order to empower women, minorities, and other marginalized people.

Constant changes of digital technologies challenge people with the need to master technological and social competences. Digital literacy includes, among other factors, the ability to learn, work, collaborate with others, and solve problems effectively in online learning and work environments, as well as handling interpersonal and group interactions in technology-mediated social involvement environments. The current study contributes to this area in exploring some effects of digital technologies, in interaction with participant personality, on online behavior in the context of digital literacy. Following our research, it seems to be evident that the impact of personality in interaction with media characteristics and communication content affect both participants’ readiness as well as actual accomplishments in different modes of communication. By conducting this research in the natural Internet environment it has improved the ecological validity of the study and enlarged the possibility for generalization of its findings. While this study has implications concerning the use of synchronous online environments and differentiates the impact of text chat from voice chat, further investigative studies are needed to examine the growing use and possibilities of synchronous communication modes.

References


Ng, K. C. (2007). Replacing face-to-face tutorials by synchronous online technologies: Challenges and pedagogical implications. The International Review of Research in Open and Distance Learning, 8(1), 1-15.


Internet Abuse among Teenagers and Its Relations to Internet Usage Patterns and Demographics

Suzan Lema Gencer and Mustafa Koc
Department of Educational Technology, Suleyman Demirel University, 32260, Isparta, Turkey // szntamer@gmail.com // mustafakoc@sdu.edu.tr

ABSTRACT
This study focused on exploring Internet abuse among teenagers and its relations to some Internet usage patterns and demographic characteristics in a digitalizing country, Turkey. It was designed as a cross-sectional research on three types of school that differ in their academic performances. The results were collected from 1380 high school students through a paper-based questionnaire. The results identified a small portion of students as Internet abusers experiencing severe problems and one fourth as possible abusers experiencing occasional problems in their lives. Excessive use, tolerance, preoccupation with the Internet, and using the Internet to escape from negative feelings were the most frequently reported symptoms of disturbed patterns of online behaviors. One-way between-groups ANOVA tests revealed that Internet abuse differed significantly based on gender and perceived academic achievement with small effect sizes, and frequency of Internet use, dominant place of Internet use and dominant purpose for Internet use with medium and large effect sizes. On the contrary, no significant differences were found based on perceived socio-economic status and the type of school attended.

Keywords
Internet abuse, High school students, Usage patterns, Demographics, Turkey

Introduction
Internet access and usage in the world has been proliferating year by year, with approximately 1.11 billion users in 2007, 1.67 billion in 2009, and 1.97 billion in 2010 (Miniwatts Marketing Group, 2010), indicating an upward trend in the number of digitally literate people. Such a rapid growth has been interacted with people’s needs and motivation. Information, communication, and entertainment have been prominent motives behind the Internet use. The advantages of the Internet are undeniable and well-evidenced in the literature. Nevertheless, excessive or unregulated usage has been associated with a condition of Internet-related disturbances which Morahan–Martin (2008) calls “Internet abuse” referring to the “patterns of using the Internet that result in disturbances in a person’s life but does not imply a specific disease process or addictive behavior” (p. 34). Some scholars or clinicians prefer to use the term “Internet addiction” to define this condition as a form of impulse control disorder (e.g., Young, 1998). Despite various approaches to the conceptualization of the condition, which is still developing and negotiated in ongoing research, studies acknowledged its existence and reported similar symptoms: school and work-related impairments, interpersonal problems, preoccupation with using the Internet, using the Internet to improve negative moods, and serious disturbances in users’ social capitals (Morahan–Martin, 2008). The present study adopted the concept of Internet abuse because it examined the condition in a non-pathological population and operationalized the Internet behavior as a continuum from normal to problematic usage.

Nowadays, the adoption of digital technologies is known to be higher in young adolescents than adults. For example, the highest proportion of computer and Internet usage belongs to the 16–24 year old age group in Turkey (TUIK, 2010), where Internet users aged 15 and older were reported as the most engaged users in Europe in terms of time spent and the content consumed on the Internet (comScore, 2009). Being the mainstream consumers and drivers of digital contents, youngsters are at the center of a lucrative digital marketing enterprise. New products are designed essentially for appealing to emotions, habits, and values in youth culture (Montgomery, Gottlieb–Robles & Larson, 2004). This in turn makes young people early adopters and heavy users of digital technologies. Such an involvement, coupled with the psychologically sensitive developments of adolescence period, can make teenagers more susceptible to digital disturbances such as Internet abuse (Chou, Condron & Belland, 2005; Tsai & Lin, 2003). Therefore, empirical investigations of factors in problematic Internet use among teenagers have been called for to help parents, teachers, and counselors in guiding teenagers about the conscious and beneficial effects of Internet use. This study attempted to contribute to this call by exploring possible interactions between Turkish high school students’ status of Internet abuse, Internet usage patterns, and demographic characteristics.
Although prior research on overall analysis of Internet abuse is prolific, the research on socio-demographic differences in this field is rather sparse or inconclusive. Gender is one of the predominantly investigated variables seen in the literature. While some studies indicated no gender differences in Internet abuse (e.g., Ferraro, Caci, D’amico & Blasi, 2007; Lee et al., 2007; Soule, Shall & Kleen, 2003), others revealed that males were more likely to become Internet abusers than females (e.g., Chou et al., 2005; Morahan–Martin & Schumacher, 2000; Yang & Tung, 2007). Although socio-economic status (SES) is an important factor especially in the developing countries, where digital divide still exists to some degree, not much attention has been devoted to this demographic variable. Only a few studies concluded that SES was not significantly associated with Internet abuse (e.g., Cakir–Balta & Horzum, 2008; Yoo et al., 2004). More systematic research is needed to clarify the role of gender and SES in problematic Internet use. Furthermore, increased availability of Internet access in educational environments has raised the recent question of whether disturbed Internet usage can be related to some academic variables. For example, the level of Internet abuse may be dependent on the schools that differ in terms of student motivation and academic achievement. School culture may also contribute to online behaviors. Perhaps, teachers, managers, or schoolmates socially encourage and reward students’ Internet usage. Empirical research is needed to examine school-related variables. Bayraktar and Gun (2007) found that secondary and high school students’ grade point averages decreased as their disturbed use of the Internet increased. Yang and Tung (2007) found that the likelihood of Internet abuse was higher for vocational high school students compared to other high school students. However, such research is very limited to make conclusive interpretations.

On the other hand, a growing body of research has investigated what kinds of Internet use are associated with Internet abuse. In particular, the amount of Internet use has been persistently explored in the literature. While it has been found to be positively associated with Internet abuse, the findings differentiated in terms of average time spent online by Internet abusers (Cakir–Balta & Horzum, 2008; Chou & Hsiao, 2000; Morahan–Martin & Schumacher, 2000; Young, 1998). Therefore, further research with different measures of Internet use (e.g., frequency) is suggested to solidify the validity of previous findings. Moreover, there are other key usage patterns that may interact with disturbed Internet use. One of them could be the type of activities engaged on the Internet. Prior research indicated that Internet abusers often preferred interactive online activities such as chatting and playing online games (Chen, Tarn & Han, 2004; Young 1998). Such research should continue in different population and time periods because online behaviors may change according to both user characteristics (e.g., age, occupation, cultural identity) and new applications that will be added on the Internet. Yet another usage pattern is the place of Internet use. Young users usually share Internet access with others at their home and schools. This may lead to a limited amount of usage. Having watched by family members, teachers, and peers can also make youngsters feel uncomfortable and thus seek alternative places. There are some commercial places (e.g., cybercafes) that provide somewhat unrestricted Internet access at a reasonable cost. Such places also function as a popular entertainment and socialization outlet for youngsters because they can gather and use the Internet together (e.g., playing multi-user games). Research demonstrated that cybercafes covered 20% of recent Internet usage in Turkey (TUIK, 2010) and more than half of their clients were teenagers (Gurol & Sevindik, 2007). Another study on Turkish students aged 12 to 18 showed that most (72.4%) went to cybercafes and frequent visitors were more likely to be problematic Internet users than rare visitors (Tahiroglu, Celik, Uzel, Ozcan, & Avei, 2008). Although limited number of studies indicated home usage as an important factor in Internet abuse (e.g., Johansson & Gotestam, 2004), further research is needed to determine the associations of various places with Internet usage.

**Purpose of the study**

The purpose of this cross-sectional study was to explore the status of Internet abuse among 15–18 years old Turkish high school students and its relationship to their Internet use patterns and demographic characteristics. As indicated in the previous section, similar studies exist in the literature, but they are inconclusive and most focus on the developed and highly-digitalized countries that may differ in certain respects. Beard (2005) maintains that disturbed Internet use may differ by a number of socio-cultural factors including, but not limited to, family conflict, peer modeling, expectations of the society, Internet penetration rate, lack of certain social skills, cultural orientation, and value beliefs about the Internet. More research is needed to examine which variables are related to Internet abuse. Therefore, the present study attempts to enhance related literature by presenting data about some rarely examined usage variables (e.g., place of Internet use), beyond the usual demographics (e.g., school type), and from a culturally distinct developing country, Turkey. The study was guided by the following research questions:
1. How do Turkish high school students’ Internet usage patterns (i.e., frequency of use, dominant place of use, and dominant purpose of use) distribute?
2. What is the status of Internet abuse among these students?
3. Are there differences in Internet abuse based on demographics (i.e., gender, perceived SES, school type, and perceived academic achievement)?
4. Are there differences in Internet abuse and Internet usage patterns?

Research methodology

Participants

Cluster sampling was employed to recruit the participants from the population of high school students in 20 schools in the city of Isparta, Turkey. Since school type was used as a grouping variable in the study, the schools were clustered into three groups: vocational (n=8), general (n=4), and Anatolian schools (n=8). This order also represents the ascending order of average students’ motivation and achievement in these schools. In accordance with the research budget and time constraints, a total of five schools were proportionally (i.e., 40%, 20%, and 40% respectively) and randomly selected from the three groups. Therefore, the final sample was made up of 1380 students who were accessible and volunteer at the time of the data collection. Demographic characteristics of the sample are given in the following results section.

Procedures and instruments

The data were gathered through a paper–based questionnaire towards the end of the fall semester in 2010. In order to ensure the content validity of the instrument, all questions were derived from the relevant literature. Moreover, the questionnaire was reviewed by two educators to ensure the face validity. The approval for the survey administration was granted by the Provincial Office of National Education in the city of Isparta. The first author visited schools and conducted the survey in two weeks. Students were not offered any incentives and their anonymity was assured by not requesting any identifying information.

The questionnaire form comprised three main sections. The first section contained several multiple–choice items that asked students to provide the following background information: gender (male, female), type of the school they were attending (vocational, general, and Anatolian), perceived SES of their family (low, middle, and high class), and perceived academic achievement (poor, average, and good).

In the second section, students were asked to provide information about their Internet usage. The frequency of Internet usage was measured by a four–point Likert scale item ranging from “not at all” to “everyday”. The dominant place of the usage was measured through a multiple–choice item with single response and “other” option that asked students to indicate the place where they most frequently access to the Internet. The related literature revealed four main motives for Internet use as: information (e.g., searching for information, doing assignments, reading news), communication (e.g., using e–mail, instant messaging), entertainment (e.g., playing games, watching videos), and business (e.g., online shopping, trading) (Koc & Ferneding, 2007). Hence, the dominant purpose of usage was operationalized on a multiple–choice item with these four options asking students to select only the one for which they most frequently use the Internet.

The last section contained “Internet Addiction Test” (IAT) that was developed by Young (1998) as one of the first standardized tests for the assessment of disturbed Internet use. Although there are several instruments available in the literature, the IAT was selected for this study because it had already been adapted into Turkish by Bayraktar (2001). It was also shown to produce valid and reliable measures in several studies on Turkish youth (e.g., Bayraktar, 2001; Cakir–Balta & Horzum, 2008). The test contains 20 items which ask respondents to rate how often they show such symptoms of damaging Internet usage as excessive time spent online, neglect of daily routine tasks, disruption of academic or job performance, concealment of online time and behaviors from others, loss of sleep, social isolation, depressive feelings if usage is restricted, and failure attempts to cut down Internet use (Table 2). It employs a six–point Likert scale ranging from “0=never” to “5=always.” Thus, the possible total score for each respondent could range from 0 to 100, with higher scores indicating greater problems associated with Internet usage. In accordance with Young’s (1998) grouping criteria based on the normative cut–off scores, the present study classified students...
with a score of 39 and below as “average user”, those with a score of 40 to 69 as “possible Internet abuser”, and those with a score of 70 and above as “Internet abuser”.

Prior psychometric studies of Internet abuse have indicated that the IAT is a valid and reliable instrument (e.g., Widyanto & McMurran, 2004). Nevertheless, a confirmatory factor analysis using structural equation modeling on LISREL 8.80 software program was conducted to verify the single–factor structure of the IAT for the present data set. This analysis was based on a covariance matrix and maximum likelihood estimation. Since chi–square value is known to be inflated and significant with large sample size, other goodness–of–fit indices including Comparative Fit Index (CFI), Normed Fit Index (NFI), Goodness of Fit Index (GFI), Root Mean Square Error of Approximation (RMSEA) and Root Mean Square Residual (RMR), which are less sensitive to sample size, were used to assess the fitness of the single–factor model. The results revealed that standardized factor loadings of the items were all meaningful in size, ranging from .43 to .69. The t–test values for these loadings ranged from 15.73 to 27.44 and were statistically significant (p<.01), indicating that Internet abuse significantly explained all items of the IAT. The goodness–of–fit measures (CFI=.96, NFI=.96, GFI=.91, RMSEA=.072, RMR=.055) indicated that single–factor model had an acceptable fit for the present data. Furthermore, the reliability of the model was checked by computing Cronbach alpha internal consistency coefficient. The coefficient was found .91, which indicated that the IAT produced very reliable measures.

Data analysis

The questionnaires were initially coded in order to quantify the items. Next, all the data were entered to Statistical Package for the Social Sciences (SPSS 17). Then, they were subjected to frequency and percentage analyses in order to check for possible errors made during the data entrance. This also helped to describe the variables. Measures of central tendency and variability were calculated for the scores obtained from IAT items. Finally, a series of one–way between–groups analysis of variance (ANOVA) with post hoc tests were conducted to find out whether there were significant differences in the mean scores of total IAT scores (dependent variable) across the levels of demographical and Internet usage variables (independent variables) included in the study. Partial Eta–squared (\( \eta^2 \)) and Cohen’s \( d \) statistics were also calculated to estimate effect sizes (i.e., strength of relationships) for the overall ANOVA tests and follow–up post–hoc comparisons respectively.

Results

Demographic Characteristics of the Sample

Due to the use of cluster sampling, the sample distribution of school type was very similar to the one in the population: 39% vocational, 23% general, and 38% Anatolian high school. Regarding gender breakdown of the participants, the proportion of females (56%) was larger than those of males (44%). When asked about the SES, the majority of the students (86%) perceived their family as middle class while the remaining evenly divided between low (7%) and high (7%) class. As far as academic achievement was concerned, students with a perception of average school performance were the largest group (64%), followed by those with a perception of good (25%) and poor (11%) school performance.

Internet usage patterns

Table 1 presents Internet usage patterns and their categories with respective percentages. In terms of the frequency of use, almost one third (30.8%) used the Internet every day, just less than half (45.5%) used it a few times a week, and one fifth (20.1%) used it a few times a month. Only 3.6% of the students reported that they had never used the Internet. Therefore, most of them (76.3%) were at least weekly users. More than half of the users (63.8%) reported “home”, one quarter (25%) reported “cybercafés,” and a small proportion (5.4%) reported “school” as their dominant place of Internet use. Moreover, a tiny proportion (5.8%) specified other places that included workplace, friends’, relatives’, and neighbors’ houses. When asked about the purpose for which they mostly use the Internet, 39.2%, 30.6%, and 29.7% indicated communication, information, and entertainment respectively. The remaining, less than one percent, reported using the Internet mostly for business purposes.
Table 1. Frequency distribution of students’ internet usage patterns

<table>
<thead>
<tr>
<th>Pattern / category</th>
<th>Frequency (f)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of Internet use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>50</td>
<td>3.6</td>
</tr>
<tr>
<td>A few times a month</td>
<td>276</td>
<td>20.1</td>
</tr>
<tr>
<td>A few times a week</td>
<td>627</td>
<td>45.5</td>
</tr>
<tr>
<td>Everyday</td>
<td>424</td>
<td>30.8</td>
</tr>
<tr>
<td>Total</td>
<td>1377</td>
<td>100</td>
</tr>
<tr>
<td>Dominant place of Internet use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>846</td>
<td>63.8</td>
</tr>
<tr>
<td>Cybercafe</td>
<td>332</td>
<td>25</td>
</tr>
<tr>
<td>School</td>
<td>72</td>
<td>5.4</td>
</tr>
<tr>
<td>Other places</td>
<td>77</td>
<td>5.8</td>
</tr>
<tr>
<td>Total</td>
<td>1327</td>
<td>100</td>
</tr>
<tr>
<td>Dominant purpose for Internet use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>406</td>
<td>30.6</td>
</tr>
<tr>
<td>Communication</td>
<td>521</td>
<td>39.2</td>
</tr>
<tr>
<td>Entertainment</td>
<td>394</td>
<td>29.7</td>
</tr>
<tr>
<td>Business</td>
<td>7</td>
<td>.5</td>
</tr>
<tr>
<td>Total</td>
<td>1328</td>
<td>100</td>
</tr>
</tbody>
</table>

The Status of Internet Abuse

The total IAT scores ranged from 20 to 100 with a mean score of 34.32 (SD=13.57). Most students (73%) were identified as average users (AUs) with a mean score of 27.56 (SD=5.32), who have control over their Internet usage. Almost one quarter (24.3%) were classified as possible Internet abusers (PIAs) with a mean score of 49.48 (SD=7.89), who experience occasional or frequent problems in their lives. The remaining fraction (2.7%) was diagnosed as Internet abusers (IAs) with a mean score of 80.16 (SD=8.92), who suffer significant problems from their usage. It can be said that a little more than one fourth (27%) appear to experience some problems with Internet use.

Table 2. Descriptive statistics for the symptoms of Internet abuse

<table>
<thead>
<tr>
<th>Symptom</th>
<th>AUs</th>
<th>PIA</th>
<th>IAs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Stay online longer than intended</td>
<td>1.89</td>
<td>.92</td>
<td>3.45</td>
</tr>
<tr>
<td>Getting impatient to go online</td>
<td>1.36</td>
<td>.73</td>
<td>2.88</td>
</tr>
<tr>
<td>Craving more time to be online</td>
<td>1.50</td>
<td>.87</td>
<td>3.25</td>
</tr>
<tr>
<td>Going online to block disturbed thoughts</td>
<td>1.76</td>
<td>.92</td>
<td>3.08</td>
</tr>
<tr>
<td>Feeling restless when not being online</td>
<td>1.16</td>
<td>.54</td>
<td>2.16</td>
</tr>
<tr>
<td>Getting complaints about time spent online</td>
<td>1.40</td>
<td>.81</td>
<td>2.89</td>
</tr>
<tr>
<td>Think of life as boring without the Internet</td>
<td>1.63</td>
<td>1.02</td>
<td>3.00</td>
</tr>
<tr>
<td>Fantasy of being online when offline</td>
<td>1.22</td>
<td>.53</td>
<td>2.32</td>
</tr>
<tr>
<td>Checking e-mails before anything else</td>
<td>1.95</td>
<td>1.19</td>
<td>3.19</td>
</tr>
<tr>
<td>Loss of sleep</td>
<td>1.19</td>
<td>.54</td>
<td>2.31</td>
</tr>
<tr>
<td>Neglect of household chores</td>
<td>1.25</td>
<td>.51</td>
<td>2.12</td>
</tr>
<tr>
<td>Getting annoyed when bothered online</td>
<td>1.50</td>
<td>.97</td>
<td>2.53</td>
</tr>
<tr>
<td>Forming new online relationships</td>
<td>1.40</td>
<td>.78</td>
<td>2.30</td>
</tr>
<tr>
<td>Negative effect on school performance</td>
<td>1.17</td>
<td>.48</td>
<td>2.01</td>
</tr>
<tr>
<td>Lowered productivity</td>
<td>1.20</td>
<td>.54</td>
<td>2.07</td>
</tr>
<tr>
<td>Withdrawing from other pleasurable activities</td>
<td>1.12</td>
<td>.42</td>
<td>1.92</td>
</tr>
<tr>
<td>Failed attempts to reduce time spent online</td>
<td>1.33</td>
<td>.76</td>
<td>2.31</td>
</tr>
<tr>
<td>Guilt about Internet usage</td>
<td>1.28</td>
<td>.69</td>
<td>2.25</td>
</tr>
<tr>
<td>Deception about time spent online</td>
<td>1.10</td>
<td>.41</td>
<td>1.85</td>
</tr>
<tr>
<td>Social estrangement</td>
<td>1.13</td>
<td>.40</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Note. AUs=Average users; PIAs=Possible Internet abusers; IAs=Internet abusers.
Table 2 demonstrates mean scores and standard deviations for the symptoms of Internet abuse measured by the IAT index for each of three groups. It prioritizes these symptoms in order of severity in the IAs. As expected due to the use of normative cut-off scores, the mean scores for each symptom increased across the groups from AUs to IAs. The most frequently reported symptoms among the IAs are related to excessive use, tolerance, preoccupation with the Internet, and using the Internet to escape from negative feelings. Similar symptoms were also reported most frequently among the PIAs. Social estrangement was the least frequently reported symptom in both PIAs and IAs.

The role of demographics in Internet abuse

Table 3 presents the number of subjects, mean scores, and standard deviations for all levels of each demographic variable. Table 4 shows one–way between–groups ANOVA results for the comparisons of IAT scores across the demographics. There was a significant difference with a small effect size in IAT scores based on gender {F(1, 1326)=30.32, p<.01, partial η²=.02}. Since the homogeneity of variance assumption was violated, Brown–Forsythe correction were applied and it confirmed the significant difference {F(1, 1135.08)=29.05, p<.01}. Male students reported significantly higher mean scores than female students. Besides, the IAT scores differed significantly based on perceived academic achievement {F(2, 1319)=22.81, p<.01}. The effect size was small (partial η²=.03). Again, due to the heterogeneity of the variances, Brown–Forsythe correction confirmed the significant difference {F(2, 384.46)=17.95, p<.01}. Follow–up post–hoc comparisons using Dunnett’s C test indicated that poor achievers had significantly higher mean scores than both average (p<.01, Cohen’s d=.42) and good achievers (p<0.01, Cohen’s d=0.58). On the other hand, there were no significant differences based on school type {F(2, 1325)=0.64, p>0.05} and perceived SES {F(2, 1324)=2.98, p>0.05}.

Table 3. Descriptive statistics for Internet abuse by demographics

<table>
<thead>
<tr>
<th>Demographic / Level</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>738</td>
<td>32.51</td>
<td>12.20</td>
</tr>
<tr>
<td>Male</td>
<td>590</td>
<td>36.59</td>
<td>14.80</td>
</tr>
<tr>
<td>Total</td>
<td>1328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>303</td>
<td>33.56</td>
<td>13.76</td>
</tr>
<tr>
<td>Vocational</td>
<td>515</td>
<td>34.63</td>
<td>13.23</td>
</tr>
<tr>
<td>Anatolian</td>
<td>510</td>
<td>34.46</td>
<td>13.80</td>
</tr>
<tr>
<td>Total</td>
<td>1328</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived socio–economic status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>79</td>
<td>32.84</td>
<td>12.80</td>
</tr>
<tr>
<td>Middle</td>
<td>1156</td>
<td>34.18</td>
<td>13.23</td>
</tr>
<tr>
<td>High</td>
<td>92</td>
<td>37.45</td>
<td>17.51</td>
</tr>
<tr>
<td>Total</td>
<td>1326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived academic achievement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>153</td>
<td>40.60</td>
<td>17.49</td>
</tr>
<tr>
<td>Average</td>
<td>850</td>
<td>34.13</td>
<td>12.87</td>
</tr>
<tr>
<td>Good</td>
<td>319</td>
<td>31.77</td>
<td>12.87</td>
</tr>
<tr>
<td>Total</td>
<td>1324</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. One–way ANOVA results for differences based on demographics

<table>
<thead>
<tr>
<th>Demographic variable</th>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F–value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Between groups</td>
<td>5461.51</td>
<td>1</td>
<td>5461.51</td>
<td>30.32*</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>238867.30</td>
<td>1326</td>
<td>180.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>244328.81</td>
<td>1327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School type</td>
<td>Between groups</td>
<td>236.42</td>
<td>2</td>
<td>118.21</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>244092.40</td>
<td>1325</td>
<td>184.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>244328.82</td>
<td>1327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived socio–economic status</td>
<td>Between groups</td>
<td>1094.76</td>
<td>2</td>
<td>547.38</td>
<td>2.98</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>243127.50</td>
<td>1324</td>
<td>183.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>244222.30</td>
<td>1326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived academic achievement</td>
<td>Between groups</td>
<td>8138.41</td>
<td>2</td>
<td>4069.21</td>
<td>22.81*</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>235353.40</td>
<td>1319</td>
<td>178.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>243491.81</td>
<td>1321</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01
The role of usage patterns in Internet abuse

Similar ANOVA tests were repeated to determine whether Internet abuse significantly differed across the levels of Internet usage patterns. Table 5 illustrates the number of subjects, mean scores, and standard deviations for all levels and Table 6 summarizes the results of ANOVA analyses.

There was a significant difference with a large effect size in IAT scores based on the frequency of Internet use \(\{F(2, 1324)=167.84, p<.01, \text{partial } \eta^2=.20\}\). Brown–Forsythe correction \(\{F(2, 904.77)=169.44, p<.01\}\) was calculated to confirm this finding because of the heterogeneity of variances. Analysis of post hoc comparisons via Dunnett’s C test revealed that the IAT scores of daily users were significantly higher than those of weekly (\(p<.01, \text{Cohen’s } d=.84\)) and monthly users (\(p<.01, \text{Cohen’s } d=1.20\)), and that the IAT scores of weekly users were significantly higher than those of monthly users (\(p<.01, \text{Cohen’s } d=.46\)).

Moreover, the IAT scores significantly differed with a small to medium effect size based on the dominant place of Internet access \(\{F(3, 1323)=22.86, p<.01, \text{partial } \eta^2=.05\}\). Due to the heterogeneity of variances, Brown–Forsythe correction confirmed this finding as well \(\{F(3, 552.06)=35.60, p<.01\}\). Dunnett’s C post hoc test showed that home access group scored significantly higher on the IAT than school (\(p<.01, \text{Cohen’s } d=.81\)), cybercafe (\(p<.01, \text{Cohen’s } d=.38\)), and other access group (\(p<.01, \text{Cohen’s } d=.56\)), and that cybercafe access group scored significantly higher on the IAT than school access group (\(p<.01, \text{Cohen’s } d=.43\)).

Finally, there was a significant difference with a medium effect size in IAT scores based on the dominant purpose of Internet use \(\{F(2, 1318)=80.15, p<.01, \text{partial } \eta^2=.11\}\). Again, due to the heterogeneity of the variances, Brown–Forsythe correction confirmed the significant difference \(\{F(2, 1144.47)=81.62, p<.01\}\). According to the Dunnett’s C post–hoc comparisons, information users reported significantly lower scores than both communication (\(p<.01, \text{Cohen’s } d=.72\)) and entertainment users (\(p<.01, \text{Cohen’s } d=.89\)). The business group was excluded from the analysis to ensure the assumptions of ANOVA because it had negligible number of subjects (.5%) and thus caused a high group size discrepancy.

<table>
<thead>
<tr>
<th>Table 5. Descriptive statistics for Internet abuse by usage patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage pattern / Level</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Frequency of usage</td>
</tr>
<tr>
<td>A few times a month</td>
</tr>
<tr>
<td>A few times a week</td>
</tr>
<tr>
<td>Everyday</td>
</tr>
<tr>
<td>Dominant place of access</td>
</tr>
<tr>
<td>Home</td>
</tr>
<tr>
<td>Cybercafe</td>
</tr>
<tr>
<td>School</td>
</tr>
<tr>
<td>Other places</td>
</tr>
<tr>
<td>Dominant purpose for usage</td>
</tr>
<tr>
<td>Information</td>
</tr>
<tr>
<td>Communication</td>
</tr>
<tr>
<td>Entertainment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6. One–way ANOVA results for the comparison of Internet abuse by usage patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet usage pattern</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Frequency of usage</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dominant place of access</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dominant purpose for usage</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*p<0.01
Discussion

Although the recent Internet penetration rate in Turkey (45%) was much smaller than those in highly–digitalized countries such as Germany (79%), the USA (77%), and South Korea (81%) (Miniwatts Marketing Group, 2010), Turkey ranked the first place among 17 European countries in terms of Internet usage with users spending an average of 32 hours online and visiting an average of 3044 web pages per month (comScore, 2009). Most of the Internet usage in Turkey also belongs to youth (TUIK, 2010). Such an involvement with the Internet warranted this investigation of whether and how prevalent Internet abuse was among Turkish high school students and what factors could have an effect on this condition.

Almost three percent of the students were diagnosed as IAs experiencing severe problems in their lives. Similar studies in the literature vary widely in terms of their rates of IAs because of using different samples, instruments, and socio–cultural contexts. After reviewing available research conducted in various countries, Young, Yue, and Ying (2011) concluded that the prevalence of Internet abuse generally ranged from 4.6% to 4.7% among adolescents, 13% to 18.4% among college students, and 6% to 15% among general population of the users. The rate found in this study is slightly lower than the general rate of adolescents. One possible reason for this might be the role of Turkey’s collectivist culture and strong family ties in youngsters’ lives. According to Hofstede’s (2001) theory on cultural dimensions, people in collectivist cultures belong to strong extended network of family, friends, and relatives that protect them throughout their lives in return for their loyalty. Such intimate connectedness can supply social and emotional needs of youngsters and left a little room for the Internet to function as a source of need satisfaction. In this study, the finding of “social estrangement” being the least frequently reported symptom in both IAs and PIAs in supports this explanation as well. Family interaction has been shown to have positive relations with the development of mental well–being and the prevention of problematic behaviors (Segrin & Flora, 2005). In fact, prior studies indicated positive effects of parenting attitudes and family cohesion on Internet abuse (Park, Kim, & Cho, 2008). Another possible explanation might be the fact that typical high school students in Turkey have a structured and busy life. Since there is a high demand but limited admittance to higher education, the centralized and nation–wide university entrance examination becomes a central part of their lives. In order to be well–prepared for this exam, most students engage in extra academic activities (e.g., attending private tutoring centers) beyond their usual schooling tasks. Hence, being occupied with such a competitive and challenging process and a sense of purpose may have left a limited amount of time to get and stay online. A further explanation could be related to the continuing progress of Turkey’s digitalization process. Although the proportion of households with Internet access (42.9%) has been demonstrating an upward trend so far (TUIK, 2011), it is still not as pronounced as that of Internet–based societies. In this sense, some students may have limited or even no usage because they need to access to the Internet at places other than their own home. A little more than two third (64%) reported “home” as the regular place of access. The findings that 3.6% of the students never used the Internet and 20.1% used it a few times a month support this explanation as well.

On the other hand, the study showed that almost one quarter of the participants (24.3%) had symptoms of potential Internet abuse with moderate problems. This finding, coupled with the evidence of highest rates of Internet abuse among the college students in the literature (Young et al., 2011), has an alarming implication that their disturbed usage may get more severe upon their transition to college life with flexible study hours and increased access to the Internet. Those students who are not able to get any acceptance for higher education may also become high risk users due to the difficulty of being employed and its associated feelings of stress and emptiness. Longitudinal research is needed to follow the changes in the level of Internet abuse. Also, research is needed to determine Turkish people’s awareness about the possible discursive dimensions of intensive online engagement. Identification of PIAs has also implications for policy makers and educators to initiate some informative actions and warning signs that may help them change their Internet use habits. Schools can implement counseling and informative seminar programs for both students and parents about the conscious and productive use of the Internet including the risk of Internet abuse, its symptoms, potential prevention strategies, online time management, Internet filtering, parental supervision, and setting and enforcing acceptable use policies. Since the level of Internet abuse in daily users and entertainment–oriented users was found to be significantly higher than their counterparts, students should be encouraged to plan and control their amount of usage and engage with alternative entertainment activities (e.g., sports, traditional games) that do not require the use of the Internet. Since students who used the Internet mostly at home had the highest IAT score within the dominant places of access, parents should be advised to place computers in a visible room where they can easily monitor their children’s use of the Internet.
The study showed a gender difference with males having a slightly higher level of Internet abuse than females. However, despite the significance, the effect size was small. This suggests that the actual difference may not be regarded as substantial enough for theoretical and practical implications. The literature has several reports showing that males are more likely to be Internet abusers than females (Bayraktar & Gun, 2007; Frangos, Frangos, & Kiohos, 2010; Tahiroglu et al., 2008; Yang & Tung, 2007). However, the present finding can hardly be taken as support for these reports due to the small effect size. Since not all studies reported effect sizes, it is also difficult to determine whether reported gender differences are weak or strong. The small difference may be related to socio-cultural and usage factors such as (a) the double–digit gender gap in Turkey (19.6%) that favors males in terms of Internet use rates (TUIK, 2011), (b) perception of cybercafes as masculine spaces that may restrict females’ places of access (Gürol & Sevindik, 2007), (c) males’ preference to use the Internet primarily for entertainment (Bayraktar & Gun, 2007) that was found to be related to Internet abuse in this study, and (d) Turkish men’s preferences for spare time activities requiring little effort (Tezcan, 1976), which are available on the Internet. On the other hand, the small effect size may indicate that gender differences in Internet abuse are narrowing down as females get more access to the Internet. Indeed, Turkish National Statistics demonstrates that the gender gap in Internet use has been decreasing from 20.6% in 2009 to 20.1 in 2010 and 19.6 in 2011 (TUIK, 2010, 2011). Future research should confirm whether gender effect on Internet abuse diminishes while the gender gap in Internet use decreases.

Perceived academic achievement was found to have a significant effect on Internet abuse. Students with the perception of poor achievement had a higher level of Internet abuse than those with the perception of both average and good achievement. Similar to gender, the overall effect size was small. However, the effect size for the difference between poor and good achievers was medium. Thus, this finding corroborates with a few previous studies in which students’ general point averages were found to be decreased as their problematic usage intensified (Bayraktar & Gun, 2007; Frangos et al., 2010). One explanation for this difference might be that poor achievers may get stressed after a while and lose their interest in academic life. As a result, they may draw their interest to the Internet so that they can escape from this stress and satisfy the gratification that they could not be able to do in their academic endeavors. Hence, school teachers and counselors should designate those students with poor grades and take actions to improve their sense of achievement.

In contrast to gender and perceived academic achievement, school type and perceived SES were not found to be significantly related to Internet abuse. Particularly, the lack of effect by school type was a surprising result because the level of Internet abuse among the schools with higher student motivation and established culture of success (e.g., Anatolian high schools) was expected to be less than one among the other schools. However, the results indicate that a student’s individual perception of achievement may be more important than his/her school’s level of success because the former was found to be significant. On the other hand, the lack of effect with perceived SES was consistent with the limited research available in the literature (Cakir–Balta & Horzum, 2008; Yoo et al., 2004). One possible reason for this may be the narrowing gap in the use of Internet across different socio-economic classes. The decreasing cost of hardware, software, and digital subscription due to the recent competition in technology industries enable most people to afford Internet access at home. For example, the rate of household with Internet access increased from 30% in 2009 to 41.6% in 2010 and 42.9% (TUIK, 2010, 2011). Moreover, the cost of Internet access at cybercafes is quite reasonable that most students can go to these places by using their weekly school allowance. As of 2008, the Ministry of Education (MEB) in Turkey equipped all high schools and 94% of elementary schools nationwide with computer technology and Internet access (MEB, 2009). Some schools open their computer labs at the weekends so that students without home access can use the Internet at no charge. In addition to all these, the seduction of the Internet, its contributions to human life, and the social status it bring into the users (e.g., being “cool” among the peers) may sustain people’s interest in Internet use regardless of their socio-economic status.

Consistent with the general argument of Internet abuse being influenced by behavioral factors, this study indicated significant effects of Internet usage patterns on Internet abuse. Regarding the frequency of usage, the effect size was large. Students who used the Internet everyday had a higher level of Internet abuse than those who used it a few times a week or a month. Moreover, students who used the Internet a few times a week had a higher level of Internet abuse than those who used it a few times a month. It can be stated that frequent users tend to experience more Internet–related problems. Many studies investigated how the duration of being online is related to the level of Internet abuse and reported positive correlations. The daily or weekly amount of time spent online by Internet abusers was found to be larger than the time spent online by normal users (Cao & Su, 2007; Hardie & Tee, 2007; Lin & Tsai, 2002; Yang & Tung, 2007). The present study used the frequency as a measure and found that Internet abuse was the most severe among daily users.
Furthermore, the most preferred place of access had a medium effect on Internet abuse. The study indicated that the level of Internet abuse was the highest among those students who used the Internet mostly at home. This is an expected finding and consistent with prior research (Johansson & Gotestam, 2004) because home access provides anytime and unlimited access. However, this finding may raise the question of whether parental control at home is limited and effective. Not surprisingly, the level of Internet abuse was the least among the students who used the Internet mostly at school. This might be due to the fact that Internet connection at the schools is usually slow, filtered, and supervised by teachers. In fact, the effect size for the difference between home and school access was the greatest, suggesting that supervision and control can be effective for the prevention of problematic usage.

Finally, the dominant purpose for Internet usage had a significant effect on Internet abuse. The overall effect size was medium but the effect size for the difference between entertainment and information was large. Students using the Internet mostly for entertainment and/or communication had higher levels of Internet abuse than those using the Internet mostly for information. This finding supports the argument that entertainment and communication are the main motives for IAs (Bayraktar & Gun, 2007). Prior studies conducted in Turkish and other societies consistently indicated that chatting and playing games were strongly related to disturbed Internet use (Chou & Hsiao, 2000; Tahiroglu et al., 2008). This study also showed that students who used the Internet mostly for information reported the least level of Internet abuse compared to their counterparts. Similarly, Tahiroglu et al. (2008) found that general information searches and school related works were related to the lowest scores on a problematic Internet use index.

Conclusions and limitations

Beside numerous conveniences, the exponential growth of Internet usage has raised concerns pertaining to the impact of Internet use on youth’s lives. Parents and educators have become worried that their children’s Internet usage interferes with their social, psychological, and academic developments. Subsequently, such concerns have prompted researchers to examine whether and how users develop disturbed patterns of online behavior including Internet abuse.

This paper focused on exploring the condition of Internet abuse among Turkish high school students and its relationships with Internet usage patterns and demographic characteristics. The results show that the prevalence of Internet abuse in this population (2.7%) is somewhat low compared to most previous studies across the world. Comparative research across different populations is suggested to explore whether Internet abuse is interacted with some cultural factors (e.g., cultural identity, religious beliefs). One fourth of the students exhibited possible Internet abuse. Follow-up research should be conducted to monitor these students in terms of whether and how their disturbed use will be intensified in the future. Excessive use, tolerance, preoccupation with the Internet, and using the Internet to escape from disturbed feelings are the most common symptoms of Internet abuse for this sample of students.

Corroborating prior research, the study indicates that Internet abuse is dependent on gender and academic achievement with small effect sizes. It was also dependent on such Internet usage patterns as frequency, purpose, and place of Internet use with medium and large effect sizes. One significant contribution of the study is its use of cluster sampling to focus on examining Internet abuse among the schools that differ in their academic performances. The results suggest that the type of school is not related to students’ disturbed Internet usage. Since students spend a significant portion of time in their schools, the effect of other variables germane to school culture is a topic worth of future investigation. Empirical research on Internet abuse is still relatively new and much more research is needed to develop strong theoretical basis as well as effective assessment and treatment tools. Hopefully, this study will stimulate further discussion and provide a basis for further research by presenting relevant evidence from a culturally distinct and digitalizing country, Turkey.

The study has the same limitations seen in the previous studies in the literature. First of all, the study was designed as a cross-sectional research. Thus, the findings indicate relationships between Internet abuse, Internet usage patterns, and demographics without specifying one as the cause or effect of the other. Longitudinal studies are needed to determine the causality among these variables. Secondly, findings related to Internet abuse are based on self-report measures gathered from a specific city in Turkey. Therefore, the generalizability of the results is restricted to the Internet users in other populations. Thirdly, having students fill–in the questionnaires at the schools might influence
their disclosure of information related Internet usage. Finally, the IAT scores represent the level of Internet abuse at the time of data collection and do not provide information about its development process.

Acknowledgements

This research was supported by grants from the Center for Scientific Research at the Suleyman Demirel University under the Project 2277–YL–10. The authors also thank to Isparta Provincial Office of National Education and participating schools for their support in providing the subjects, and to Gulin Ulusoy and Melike Sultan Yasar for their proofreading of this paper.

References


Digital Assessment Literacy — the Core Role of the Teacher in a Digital Environment

Liat Eyal

Rahavat-Ilan 4, Ramat-Ilan, Givat-Shmuel 54056, Israel // eyaliat@gmail.com

ABSTRACT

One of the main functions of a teacher in a digital environment is student assessment. The need for assessment literacy based on measurement and quantitative data is weakening, both in terms of the traditional approach of the assessment on which it is based on, and given that information technologies can address these needs effectively. The assessment literacy required of a teacher today is of a completely different kind — one that is adapted to the digital environment and tailored for the pedagogical approaches of the 21st century. This article will focus on the skills, abilities, and perceptions required of the teacher in the digital environment with respect to assessment, and will demonstrate the importance of adapting the various technologies to the different assessment purposes. This definition of the term “Digital assessment literacy” is based on a doctoral thesis that examined the relationship between the technological environment and the teaching, learning and assessment processes in online courses (Eyal, 2010).

Keywords
Literacy, Digital assessment literacy, Teacher role, Digital environment

Literacy and assessment literacy

The term “literacy” has diverse interpretations (Wagner & Kozma, 2003). Common to all is a view of the level of literacy as a measure of the quality of human capital of a society or a particular area. Literacy develops in interaction with the environment (Vygotsky, 1987). I will argue that the environment is not only the human environment, but also the digital environment. In recent years, the term "assessment literacy" appears in the literature in two senses: One sense refers to the collection of the teacher's skills in test development, the composition of closed questions, the development of assessment rubrics, and statistical analysis of cumulative data for variety of teaching and learning needs (Donoho, 2000; Popham, 2004). In an article published by Popham (2004), the lack of assessment literacy was presented as “professional suicide”. Popham claimed that experts’ achievements in various fields are measured based on external measures forced upon them. Teachers are also measured by their students’ performances in tests, but surprisingly, they usually ignore this indicator as a measure of the quality of their instruction. This same assessment literacy, discussed by Popham, is the ability of the teacher to significantly delve into and interpret the test results. Alternatively, Stiggins (2002) defined an “assessment literate teacher,” as one who knows what assessment methods to use to collect information on the students’ achievements, conducts a dialogue about effective assessment results, using the ranking scores, reports, and portfolio, and understands how to use assessment to increase the motivation of learners and include them in the learning process. However, neither definition addresses the abundance of opportunities that the digital environment provides and the variety of skills required of a teacher and of learners to act in such an environment (Eshet-Alkalai, 2004).

In this article I argue that: (a) the teachers need for assessment literacy based on quantitative data measurement is disappearing, partly because it is based on traditional approaches of evaluation and because today’s advanced technological tools fully support teacher's work; (b) teachers must have assessment literacy, but for an entirely different kind of assessment—one that incorporates the skills mentioned in the definition given by Stiggins (2002) and tailors them to the digital environment; and (c) as part of the teachers’ role as evaluators in the 21st century, they should also know when to relinquish this responsibility to the students, in order to develop self-regulated and reflective learners.

This article seeks to highlight a different aspect of assessment literacy—digital assessment literacy. This term has not yet appeared in the academic literature, and refers to the role of the teacher as an assessor in a technology-rich environment.
Learning Management Systems (LMS) benefits and limitations

Monitoring students’ learning progress has always been required as part of any teaching and learning approach. "Assessment" is a term that includes various methods and ways of gathering information on the nature of the learners’ performance—the learning process and progress in studies (Birnbaum, 2000; Schank, 2001). Effective assessment includes both systematic and non-systematic collection of any of information that may contribute to understanding the learners’ place in terms of knowledge acquisition. The process involves analyzing and interpreting the information, as well as making judgments based on that interpretation.

As analysis, interpretation and judgment are processes that may be affected by external and subjective factors; some suggest the use of measurement. Measurement allows quantitative description of a particular characterization of an individual, expressed in numbers. Some argue that the combination of assessment and measurement provides a thorough and accurate picture, based upon which practical conclusions can be drawn (Wagner, 1997). A test is a systematic process in which an aspect of student behavior is quantitatively evaluated (Suen & Parkes, 2002). For several decades this system of assessment has been criticized for a variety of reasons, including the separation between the teaching-learning process and the evaluation process, the relatively low level of thinking required, and the quantitative reporting of results, which does not contribute to students’ progress. In the last decade, the central argument against the tests system is that their predictability is limited to the field and context in which the students are tested, and that they do not predict student problem solving ability, teamwork, good work habits and honesty.

However, this is still the most common way of measuring the achievements of learners in education. In the last decade, the introduction of learning management systems (LMS) has helped streamline the testing assessment process. An LMS allows teachers to develop assessment items (in this case, a test), assign them to students, receive their computerized answers, and edit different segments (e.g., all students’ answers to question X) to help identify specific weaknesses in students and manage feedback. In other words, the system may provide teachers with tools for efficient management of the evaluation process as whole, including the management and organization involved in carrying out tests (Paulsen & Keeagan, 2002; Hall, 2001; Greenberg, 2002; Liu, El Saddik & Georganas, 2003; Wang et al., 2004). Familiarity with Learning Management Systems is a basic level of Digital Assessment Literacy needed by teachers in the 21st century.

Concentration of assessment data and cumulative documentation in computerized systems benefits all parties involved in the educational processes. First, the learners benefit because they have information about their scores, their implementation of tasks on time and their attendance records, as well as an overall picture of their learning situation relative to other students. The teachers benefit because it is possible to address a variety of learning styles and levels, and strengthen the personal connection between teachers and learners, thanks to the potential for ongoing dialogue and personal feedback (Eyal, 2010). Principals may also use the digitally displayed assessment data; they can receive a general profile of a single student or class profiles at different levels of comparison. Furthermore, parents’ access to data on their child’s achievements makes them aware of vulnerable points where improvement is required and encourages parental involvement on a personal level.

Many teachers mistakenly believe that repeating lessons will improve students’ achievements. But, in fact, computerized documentation of assessment data is an ideal means for learning. Use of the documentation is immediate and can provide a complete and comprehensive picture of performance as a basis for planning and improvement. Collection is easy and fast, the information is available, simple and efficient, and do not take up teaching time or interfere with the class. Computerized documentation of collection data enables precise assessment, reflection, and feedback. It helps in remembering activities and events, allowing reuse over time. Computerization of tests streams the data into the system and teachers have only to decide how and when to use the data, and most importantly, to perform any required intervention in a timely manner (Globman & Kula, 2005; Dede et al., 2002; Smith, 2006). Smith (2006, p. 1) quoted Tim Wagner, Director of Educational Technology, the U.S. Department of Education of education technologies of U.S.: "These systems often allow them to analyze data in real time, so they can solve problems as soon as they arise." In other words, an LMS also can be used as a diagnostic tool that strengthens the capabilities of assessment on the one hand, and learning on the other (He & Tymms, 2005). It is important to emphasize that information systems do not replace the teachers’ role in providing grades, composing knowledge questions, and deciding who needs help; these roles are still performed by teachers, with the benefit of the information provided by an LMS. A study that examined the use of LMS in large classes (Eyal, 2010) found that the load on the teachers was reduced owing to teachers’ efficient use of the system, allowing them additional time for...
planning, developing and updating custom assessment items, reviewing and monitoring performance of students, giving detailed personal feedback, and searching for additional sources of information.

There are several secondary benefits to the use of an LMS: The teacher is the primary "evaluation authority." He is usually the measurement tools designer; he evaluates and draws conclusion based on the results. Typically, he performs all these actions alone. The use of computerized tests can work to the mutual benefit of a team of teachers who can build a bank of assessment items from which different tests can be assembled each time, according to the unique needs of each student or group (Bennett, 2001). The shared work effort lightens the teachers’ workload, and enables the data to be saved and easily retrieved whenever needed, and over time. In addition, an LMS enables the tracking of learning behavior (Mor, Minguillon & Carbo, 2006). Most such systems incorporate mechanisms for documenting information about students, such as how frequently they log onto the course site, the history of visits, number of posts, length of stay, and so on. Data Mining draws a rich picture of the learner, which can be used according to the learning patterns and needs of each learner (Fichter, 2003).

One of the criticisms of multiple-choice tests is that they are based on constructed problems with algorithmic solutions that encourage intentional learning of correct answers instead of higher-order thinking, which reduces the scope of the evaluation (Bennet, 1993; Osterlind, 1998; Resnick & Resnick, 1992). However, computerized assessment items may include graphics, sound, animation, and multimedia with response options at different levels. Scalise and Gifford (2006) offer a taxonomy of computer-based questions and tasks, organized into seven categories of computerized interactive questions from a series of multiple-choice questions (selection, prioritization, completion) to questions that require the examinee to complete knowledge construction and interpretive level response. The media may create response scenarios that cannot be perceived when taking a paper and pencil test, for example, clicking on an area where an image is displayed graphically (chart or map), interactions that play sounds, or analysis of animations or clips (Eyal, 2010; Parshall et al., 2002).

The risks of using tests also should be taken into consideration, for example, security issues. In practice, it is not always possible to monitor the tests taken by students at home. We cannot ask them where they have placed their study materials, nor is it feasible to ask them not to be in contact with their classmates during the exam (Suen & Parkes, 2002; Petty, Johnston & Shafer, 2004). Beyond the security constraints and technological points of failure, there are also pedagogical risks (Beichner, 2006; Hamilton, Klein & Lorie, 2000). Instead of serving as a catalyst for the integration of new pedagogy, based on social constructivist approaches (Vygotsky, 1978), the ease of use and time saving nature of technology sometimes tempts teachers to overuse them. Additional risks may be overuse of multiple-choice questions as a primary means of assessment or abandonment of aspects of self-learning, due to dependence on the teacher and insufficient collaborative learning, or due to the inability to see their colleagues' products (Eyal, 2010).

**Constructivist & creative use of traditional tools**

Let's take the use of computerized tests a few more steps forward. Interactive technologies provide rich sources, extending the learning environment and opening up a world of possibilities in planning instruction and assessment (Comeaux, 2005). If used creatively, multiple-choice tests can promote constructivist learning (Scalise & Gifford, 2006). For example, they may be useful for self-diagnosis of the learner's abilities in various fields, research, diagnostic assessment of learners for the design of complex pedagogical processes, practice, and feedback (Eyal, 2010). Such advanced and creative use of the LMS indicates the teacher has a higher level of Digital Assessment Literacy, reflected in the integration of new pedagogical concepts regarding assessment together with technological tools.

Until now we have analyzed the potential of computerized documentation and organization of assessment and behavioral data. From the traditional perspective, the teacher sets the criteria for evaluation, plans assignments for students, collects information about the learning, and uses it to improve learning outcomes and plan future instruction. This is the most common approach to teaching and assessment, which is used for the purpose of teaching basic skills. To evaluate how well the goals were achieved, objective measurement methods are employed (Black, et al., 2004). Sometimes, the added value of learning is measured in order to guide the learner. The teacher checks the degree of content mastery: (a) concepts, principles, and prior knowledge; (b) mastery of procedural knowledge (following instructions, operating appliances, basic thinking skills, and knowledge of sources and how to access them); (c) control of expression; (d) handling of maps, graphs; and more. The teacher must provide opportunities for
practice and repetition with feedback, demonstrations, practice of skills to mastery, observation and emulation of activities, and sorting and memory tasks (Birnbaum, 2000; Globman & Kula, 2005). The teacher’s role as an assessor in these teaching activities is dominant. Computerization of these processes is highly efficient, and the teacher needs to recognize the range of potential options in this area and make skillful use of them. Digital assessment literate teachers also should be aware of the technological, pedagogical, and ethical limitations, and make decisions regarding activities accordingly.

**Advanced assessment methods in a digital environment**

Learning methods in a 21st century digital environment, such as self-directed learning, collaborative learning (Garrison & Anderson, 2003), place the learner and the community at the center of the learning process, while recognizing the differences between learners, with the rate of learning adapted to the individual and his abilities, preferences and needs. Learning in a digital environment develops the learners’ writing and overall expression skills, and enables high-level discussions, exposure to rich and up-to-date materials, as well as immediate feedback (Bonk, et al., 2000). In addition, this learning method makes it possible to equip students with the cognitive skills necessary for the information age, and skills that are vital for coping with the vast quantities of information, for example: problem solving, critical thinking, creativity, self-learning strategies, meta-cognition, reflective thinking, social discussion skills, team work, and personal skills, such as persistence, curiosity and initiative (Passig, 2000). Eshet-Alkalai (2004) offered a detailed conceptual framework for the term 'digital literacy' that includes: photo-visual thinking; reproduction thinking; branching thinking; information thinking; and socio-emotional thinking.

Unique characteristics of the digital learning environment require appropriate assessment (Liang & Creasy, 2004; Petty, Johnston & Shafer, 2004), or the result is a paradox. Salomon (2000) called this an "evaluation paradox": If all components of the environment (pedagogy) remain unchanged, then the computer-assisted learning outcomes will be insignificant. By a similar argument, according to researchers and educators, the best way to evaluate the digital learning environment is by focusing more on the learning process the products and less on the use of tests (Huba & Freed, 1999; Petty, Johnston & Shafer, 2004). It can be maintained that when it comes to the advancement of learning and thinking and the improvement of teaching, there is consensus among the researchers regarding the centrality of assessment in general and formative assessment in particular (Black & William, 1998; Shepard, 2000). For this reason, the application of alternative assessment methods is recommended in addition to traditional methods (preferably in place) in order to produce rich evaluative information on students and a more comprehensive picture of their achievements (Birnbaum, 2000). Evaluation culture that emphasizes assessment as part of the instruction-learning process provides nourishment and guidance, as well as evidence of students’ thinking, learning abilities, and master of materials. This is a formative evaluation, also known as assessment for learning. This is the process of searching for evidence used by learners and teachers to determine students’ progress in their studies, and set objectives and how to best accomplish them (Assessment Reform Group, 2002). According to this approach, teachers determines the objectives, designs the tasks and determines criteria for assessing performance, evaluates the students, and produces constructive feedback that clarifies the strengths and how they can be developed, as well as the points that require strengthening (Birnbaum, 1997). The importance of learning increases when students participate in the entire process.

Alternative assessment methods in a digital environment offer students a rich variety of learning methods that enable them to exhibit higher levels of thinking. They are able to present more complex knowledge in the context of everyday life and develop a deeper dialogue about the learning content (Comeaux, 2005). Technology enhances the teaching and assessment capabilities, and creates opportunities for improvement and diversification in the evaluation of learners, including addressing written communication skills, cooperation, teamwork, and reflective thinking (Eyal, 2010; Liang & Crazy, 2004). In addition, digital environments can provide a solution for the diversity of learners, who are assigned assessment tasks and learn at a pace adapted to their needs (Alderson, 2006). Alternative assessment options for digital environments can include the following:

**Online tasks**

The overall assessment approach recommended in the digital environment is to forego large intensive tasks, the results of which we usually only see at the end of the process. This type of pedagogical approach to assessment, which includes the integration of technological tools, constitutes a cornerstone of the teacher’s digital assessment
literacy. Instead of summary assessment tasks, there is a "dismantling" of the learning process into short performance tasks, including indicators, is compatible with the digital learning environment and strengthens the link between learning goals and deliverables (Eyal, 2010). Digital environments can serve as platforms for various tasks including the solving of complex problems that require the locating information from a variety of sources, creating instructional videos on various topics, team assignments, collaborative writing processes, research tasks, and projects that ultimately yield a product. A Content Management System (CMS) can serve both as a learning environment and as the development and publishing platform for tasks with various components (such as formatted text, photo gallery, surveys, feedback, TalkBack), with products in different formats and a link to environments such as the Web 2.0 platform. These platforms invite the development of new criteria for evaluating learning, such as originality, quality, and peer reactions to the product (Eyal, 2010). The following are some examples of how to combine short-term implementation tasks with technological tools, as methods of alternative assessment:

**Digital portfolio**

The basic concept of using computerized portfolios is not new to education systems, but Web 2.0 technologies have increased the involvement of learners in the evaluation and collection of data. Web 2.0 and other technology tools are making it quicker and easier than ever to create digital portfolios of student work—a method of showcasing student progress that experts say increases student engagement; promotes a continuing conversation about learning between teachers, students, and other participants in the learning process; and extends academic lessons outside the classroom (Ash, 2011). The digital portfolio may include written, recorded, or visual items, homework, documentation of processes, tests, performance tasks, deliverables, and more. The portfolio is the ideal way to evaluate learning in a digital environment, mainly because students are partners in determining the content (Suen & Parkes, 2002). Use of multimedia allows the publication of work to a broad audience, as opposed to the traditional presentation to the teacher only. The portfolio structure and content can be easily adjusted to varying needs. You can search within the portfolio, and secure the entrance to the digital portfolio. Teachers with a high level of digital assessment literacy might opt to use the digital portfolio to share evaluation components using community interaction tools such as discussions or construction of sub-communities. Finally, it is proposed to use the digital portfolio to assess students’ thinking, especially as a tool that helps students reflect on the learning processes (Hill, 2002).

**Forums**

Although the use of forums for learning today is perceived as relatively outdated, there are learning environments that include them as part of the learning process (Lieblein, 2000). The use of forums might move on a continuum: for example, from technical support only, through to a central role in the publishing of tasks and work products. Researchers have put forth different methods for the analysis and assessment of activity in the discussion groups: Some address the quantitative dimension—the number of times a learner posts in the discussion group (Tirosh, 1999), and some focus on the structural and qualitative dimensions of discourse and social interactions in the discussions (Fahy et al., 2001; Garrison et al., 2001; Henri, 1992; Spatariu et al., 2004).

In a comprehensive study in this area, the use of forms was analyzed as a whole, with reference to structure and content (Fahy, Crawford & Ally, 2001). The analysis identified five patterns of content: questions, statements, quotations, responses and links. In order to examine the components related to social networking discussion groups, a Transcript Analysis Tool (TAT) was proposed, which allows measurement and evaluation of the structure and content, focusing on the patterns of the exchange on the network. The researchers hypothesized that a holistic analysis relating to the communication within a discussion group as a “unit” would provide a richer set, and thus maintained that the discussion group provides "a goldmine of information on psycho-social dynamics" (Henri, 1992). The study examined a group discussion that was part of an online course. The structural characteristics studied include "physical dimensions" of the network, and levels of social interaction. The model included dimensions of both structural analysis (e.g., group size and potential contacts) and content analysis (e.g., question types, messages that include personal information, and so on) in the discussion group. Today, most of the activity taking place during learning through discussion groups has been replicated in the social networks.
Online peer assessment

Peer assessment can focus on outcomes or on process, and is characterized by dimensions related to objectives, content and how the activity is managed (Topping, 1998). Characterization and presentation of the evaluation criteria are key components of this assessment (Miller, 2003). Many studies report peer review as an effective strategy to empower the learning process (Falchikov, 2003; Topping, 1998). Peer assessment helps learners improve their products by developing a deep understanding of the assessment criteria and their significance, providing opportunities for "learning by example" (Ronen & Langley, 2004) and from classmates’ feedback. It turns out that the evaluation itself is as important as the feedback received by students from their peers (Dominick, Reilly & McGourty, 1997; Zariski, 1996). Many studies compared peer assessments to teachers’ evaluations (Falchikov & Goldfinch, 2000). In practice, there is disagreement as to the legitimacy of using peer scores; therefore it has been proposed that peer scores undergo additional stages of validation based on comparison with teachers’ assessments (e.g., McGourty & Reilly, 1997). Based on these studies, it was concluded that peer assessment should not be used as a substitute for teachers’ assessment nor in addition to it, and that this method degrades the quality of assessment and its purpose on the whole. However, peer assessment involves a great investment on the part of the learner and therefore deserves a suitable reward, to reflect the true quality of assessment and not just the fact of its execution. A deep examination of peer assessment is a great deal of work for teachers, and it almost impossible to provide immediate feedback on its quality. On the other hand, a lack of feedback reduces the motivation to perform a quality assessment. The main obstacle in the implementation of peer assessment is the organizational and management skills required of teachers; here too, the use of technological systems may help in the effective organization and presentation of information and provision of feedback (Liu et al, 2002; Cuddy et al., 2001; Davies, 2000).

The above assessment processes increase involvement of the learner in assessment while decreasing the weight of the teacher’s assessment. Teachers with digital assessment literacy share the information on the criteria and their establishment; creates, adapts and assigns tasks to learners; and together they collect information about learning, relying on information documented using technological tools. After the teacher documents and summarizes the information, he discusses the results with students, and together they interpret them. The conclusions are used for planning future teaching and contribute to improved definition of tasks and student empowerment (Eyal, 2010).

The need for the teacher to step aside

In advanced learning processes, especially in the digital environment, part of the teacher’s role as an assessor is to know when to step aside. The skills required of learners to successfully function in the 21st century include: locating and acquiring knowledge independently; wise use of knowledge to solve problems; informed choice and critical evaluation, at the same time developing standards, and communication and collaboration skills. In general, it can be said that in order to function successfully in the 21st century, a person must be capable of adaptation and autonomous thinking; in other words—a capacity for self-directed learning that will persist throughout life. This means that students themselves lead the learning process: Learners will determine the objectives, choose the ways to achieve them and develop their own indicators for evaluation. Thus, students will develop their awareness and understanding of the learning process they undergo (Black & Wiliam, 1998). This meta-cognitive awareness contributes both to improve achievement and foster self-learning skills. The term "self-directed learning" refers to students’ ability to consciously monitor their thoughts, feelings and behavior while learning (Zimmerman, 1986, p. 307). This means that the self-regulated learner can channel and focus his thoughts, his feelings and actions in order to achieve the learning goals (Pintrich, 2000; Zimmerman, 2001). All self-directed learning theories are based on the premise that learning is not something that happens to learners, but something that the students do. This understanding change the educational focus and emphasizes the personal strategies they employ at their own initiative to improve educational outcomes and the learning environment. The characteristic common to all theories dealing with this issue is the self feedback "loop"—students monitor the effectiveness of learning methods and their learning strategies, and respond to feedback that they themselves produce in a variety of ways (Zimmerman, 2001).

Learning environments that foster learning and self-regulation

To realize the objectives of fostering complex self-guidance in learning, the learning environment should be flexible, allow freedom of choice (Schunk, 2000), and rely on democratic principles. The more these principles are reflected
in the learning environment, the greater the odds of developing a climate that promotes learning (Silberstein, et al., 2001). Flexibility can be accomplished through the ways in which the learning environment is organized, the time allotted to different types of activity, modes of learning and the content taught. A digital learning environment can offer a rich variety of choices and opportunities to select the time, subject matter, methods of learning and more. The learners can choose what and how to learn, what to focus on, which sources and learning materials to use, when to learn and at what pace, with whom to learn, and how to be evaluated. The multitude of opportunities and diversity of subjects contributes to inculcating norms of "behavior of choice" (Silberstein et al., 2001), while fostering a positive image and strengthening confidence, freedom the fear of making mistakes, and the courage to take chances. A digital learning environment that encourages choice may allow data collection from various sources, deployment of alternatives, comparing various options, setting priorities, making selections taking into account existing constraints (Silberstein et al., 2001). At the same time, the ability to make rational and well-founded choices is needed. Research-based learning, with the performance of tasks and information gathering, promotes freedom of choice and provides learners with guidance in their choices. Coping with these methods brings about a sense of responsibility for the selections made (Fouts, 2000; McLoughlin & Reid, 2002; Pellegrino, Chudowsky & Glaser, 2001; Quellmalz & Haertel, 2000).

I present herein some examples for learning and assessment that promote self-directed learning in digital environment and apply the principles of freedom of choice and creation of a community:

**Blogs**

In recent years, blogs have been adopted by educators as a means of teaching (Twist, 2004), stemming from the belief that their integration in the learning process increases motivation and student autonomy. Students feel greater "ownership" of knowledge; they can share with their classmates and create peer learning (Hyung, 2008). This is supported by theory of development put forth by Vygotsky (1978), who points out that the realization of the student's development potential depends on the experience of social interaction with trained personnel, an adult or friend. Students who participate in the writing of a blog as part of a digital learning environment develop personal writing skills and reflective thinking, as well as interpersonal interactions. They describe their experiences, try to express moments of inspiration, reveal things to others, and impart knowledge or life experiences. This style increases their confidence and self-image, and helps develop a positive approach to themselves as learners. The learner's blog is his personal learning space. Digital assessment literate teachers should beware of making the blog an obligating task (Hurlburt, 2008). Digital assessment literate teachers should foster students’ sense of responsibility for their own blogs, while obliging them to read their classmates blogs, and set a suitable amount of feedback and responses. The blog assimilation process is a lengthy one; sometimes the resulting benefits are apparent only at the end of the process. Teachers need to spend more time to evaluate each blog—throughout the writing process and not just at the end.

Despite the blog's contribution to teaching and learning processes, there are those who cite the problem of its use in the context of evaluation. Hurlburt (2008) claimed that she does not see any real possibility for scoring the blog. At the same time, in recent years, teachers’ deliberation on the assessment of blogs has led to different schools of thought, criteria and scales. One school of thought, based on summative assessment, for example, introduced in a blog by Megan Poore (2008), presents clear criteria that are to be met by students: content criteria such as ideas, analytical skills, creativity, and information gathering; criteria related to writing posts such as the quality of writing, relevance of links or attached media, frequency of writing, reference to the community, provision of feedback; and criteria related to design and management, such as how the blog is organized, its look and feel, and so on.

In post written by Mark Sample (2009) in his blog, he suggests a five-level scale for the assessment of students blogs: From a blog that lacks focus or tends to repeat points already raised and posted by other students to focused blogs that show in-depth thought. Another school of thought emphasizes formative assessment based on the learner’s ability to cope with research and deeper levels of learning, while emphasizing the process and not just the final product. For example, is the writing based on outside sources of information or only on independent thinking, what additional perspectives are presented in the blog. The learner is an active partner in this assessment process, providing the instructor with information on how he learns, leading to an additional criterion—whether or not it is possible to discern the author’s reflections on the issues he raises.
At an international conference on media education (ED-MEDIA) held in Canada in 2006, findings were presented on the integration of blogs in courses (Birney & Barry, 2006). The researchers noted that the results did not indicate deep reflection as a result of writing blogs; as an explanation, they raised the possibility of a fault in the assessment process because students were not required to write the blog and were not trained in reflective writing. Reference to these aspects of evaluation during the course of writing the blog constitutes advanced digital assessment literacy on the role of the teacher.

**Wiki as collaborative writing**

A wiki is a collection of linked web pages accessible for editing and shared by several people together (Leuf & Cunningham, 2001). The term "wiki" describes both the pedagogical approach and the technological tool, one example of the significance and impact of technology on pedagogy. Wikis have a variety of uses: A survey conducted in 24 universities (Schwartz et al., 2004) points to the wiki as a tool for creating interactive activities, advertising content, viewing collections of links/information, projects, FAQ collections, and even as textbooks (Ravid & Rafaeli, 2006). Learners view the wiki as a valuable learning tool. A wiki environment is essentially different from other environments where there is dialogue between learners. The basis of the wiki is the article or text, on which the activity is performed. On the other hand, discussion groups, e-mails, blogs based on an ordered or hierarchical chain of responses should be evaluated accordingly. Wiki platforms may also serve the needs for assessment of cooperative learning and are effective for measuring cooperation because they contain records of the content editing by all students on each page. (Bruns & Humphreys, 2005; Forte, Andrea & Bruckman, 2006; Meishar-Tal & Tal-Elhasid, 2008; Reitman, Augar & Zhou, 2005; Voss, 2005). The tool's main weakness is that it does not necessarily result in meaningful creation. In many cases, work on the wiki is network-like in nature, with each participant focusing on their own mission, while learning from others contributes to them — this as opposed to people working together on the same task (Siemens, 2004).

The applications and platforms presented may be integrated as assessment activities that reflect in a complete and qualitative manner the learning process in the digital environment. Combined with tools, such as self-assessment, personal practice and reflection, the foundation for more efficient self-directed learning processes is laid. Personal assessment plays a major role in the evaluation process, most of which is passed on to learners. The role of reflection is increased, and feedback can be primarily provided by the learner himself. Teachers do not intervene unless their help is required to remove obstacles to student progress. This intervention is carried out through feedback, which promotes self-esteem and learner self-feedback. In this way, assessment is interwoven into the teaching and learning processes. Assessment focuses on the students’ ability to use previously acquired knowledge to cope with new situations, as well as the ability to apply this knowledge to guide educational activities. Teachers assist and encourage students in the self-assessment of their own stated goals, and encourage them to cooperate within and contribute to society. Teachers help students choose or create content matter according to their needs. Learners need to critically produce information, adopt or develop appropriate criteria for evaluation, and establish their own self-knowledge. Each learner needs to use self and mutual feedback to evaluate the achievement of their true goals. This evaluation style is particularly suited to the information age, characterized by skepticism about knowledge: Is there an agreed-upon truth, which can be the basis for a unified program of study? Is it clear how learning occurs and how teaching promotes it? Is there a theory of learning and teaching that is better than others and accepted by all? Apparently there is no agreed upon truth, there is no "right" and "wrong" (Levin & Nevo, 2000; Wiggins, 1998), so there is no justification for judgmental feedback on the role of the teacher because there is no agreed upon criteria for what is "right."

**Conclusions**

Based on the above, several conclusions can be drawn about the role of teachers as an assessors in the digital environment: a) The role of teachers who appreciates a digital learning environment is primary and significant; b) wise use of technological tools to assess learners is essential for the students, teachers, and for other students participating in educational processes; c) teachers in the 21<sup>st</sup> century prefers to use technologies that advance the assessment methods that emphasize the learning process, enable peer assessment and develop reflective abilities.
A teacher with digital assessment literacy will be able to intelligently use various applications and technological systems to advance students, adapting a variety of assessment approaches. The level of literacy moves on a continuum, starting from the use of learning management systems as part of traditional assessment processes, such as computerized tests; to a higher literacy level that, in addition to traditional processes, includes the implementation of tests, tasks and projects in a digital environment, the performance indicators for which are determined in cooperation with students; through to implementing advanced estimation approaches based on constructivist-social learning and the development of self-targeted learning, where as part of the assessment teachers must also know how and when to delegate the processes of assessment to the students.

The following abilities and skills are required of teachers for digital assessment literacy:

**Basic Digital Assessment Literacy**
- The use of digital tools in all phases of the evaluation process: from design to drawing conclusions.
- The use of an LMS database to enable effective and focused acquisition of information about students, identifying sources of error in the teaching process, identifying difficulties on tests on both personal and class-wide levels, and examining various cross-sections of data to draw conclusions and plan instruction.
- Production of relevant assessment reports from within the LMS for various parties involved in the educational process.
- Routine testing of the effectiveness and suitability of the selected technological tools and components.
- Organization and documentation of all the mutual feedback and their efficient use to monitor and promote learning.
- Management of formative and summative assessment scoring and interpretation of results based on a digital database.
- Effective use of assessment data in the digital database for pedagogical decision making and for planning the teaching–learning–assessment processes.

**Intermediate Digital Assessment Literacy**
- Minimization of the number of computerized tests used for learning assessment.
- Ability to diagnose and assess a range of study and developmental areas using performance tasks that incorporate various technological tools.
- The use of variety of methods and digital tools to gather information on the progress of learners.
- The selection of assessment methods and technological tools appropriate to the learning objectives and teaching methods.
- The combination of several technologies for assessing and measuring learning, to increase the validity and cross-check information.
- Development of appropriate criteria for evaluating performance in a digital environment, and using feedback and guidance to promote learning.
- Awareness and ability to cope with the risks and inherent ethical issues associated with the use of digital assessment tools.

**Advanced Digital Assessment Literacy**
- Share the methods for assessment and the formation of evaluation criteria with learners, using learning systems and applications that enable transparency and collaborative writing.
- Encourage cooperative learning by having students expose their learning outcomes to their peers at all stages of implementation and at the end of the learning process (digital portfolio, personal learning environments, wikis, blogs, podcasts, publishing, multimedia databases, and so on).
- Encourage students to evaluate their peer through continuous integration of collaborative technologies that enable comment and discussion, while developing skills for learners and promoting the giving and receiving of high-quality feedback.
- Collaboration with others as part of the assessment process, by disseminating information, and providing access and permissions to various digital environments.
• Ability to produce rich assessment information about both learners and the learning-teaching process, based on documented data collection and cross-checking a variety of digital tools over time, and the ability to use interpretation to promote learning.

• Encouraging learners to use self-assessment and reflection, using advanced digital technologies such as writing a blog, computerized practice tests.

• The creation of online anonymous feedback and evaluation surveys on the learning-teaching process and the teacher, and the use of the information obtained for self-reflection to improve instruction and assessment.

• The ability to identify situations that require attention and sensitivity to the learner’s feelings, identifying the learner’s needs in the social and emotional realms, and developing his abilities through a thorough reading of the deliverables, such as personal blog.

• Providing choices for learners with regard to goals, tasks, information sources and products, according to personal preference, while opening their eyes to a variety of options offered by the Internet.

The level of digital assessment literacy is an index of teachers’ quality and professional development. Digital assessment literacy, like any literacy, develops through interaction with the environment, not only the human environment, but also the digital environment and the opportunities it provides. Teachers in the digital environment need to acquire knowledge and skills to help in the selection and use of assessment tools. To this end, the involvement and guidance of a knowledgeable mentor is needed to expose teachers to both the overt and covert potential for assessment in the digital learning environments as part of teachers training process.

References


Integration of Multimedia Courseware into ESP Instruction for Technological Purposes in Higher Technical Education

Shu-Chiao Tsai
Department of Applied Foreign Languages, National Kaohsiung University of Applied Sciences, Taiwan, R.O.C // achiiao@cc.kuas.edu.tw

(Submitted October 29, 2009; Revised January 28, 2011; Accepted April 30, 2011)

ABSTRACT
This study reports on integrating ESP (English for specific purposes) multimedia courseware for semiconductor technology into instruction of three different language programs in higher education by using it as a silent partner. It focuses primarily on techniques and tools to motivate retention of under-prepared students in an EFL setting. The courseware design was based on Mayer’s multimedia learning cognitive theory, and the language learning focus drew on Chapelle’s suggested criteria for development of multimedia CALL (computer-assisted language learning). This learner-centered instruction was compared with a traditional teacher-centered one without courseware integration. Evaluation of the instruction was based upon data from pre- and post-tests, and two questionnaires related to students’ learning satisfaction and attitude. The results suggested that students in all three programs have benefited from the courseware integration and were satisfied with practices for learning professional knowledge and English skills provided by the courseware. Students in the weekend program of recurrent education who were both older and had greater work experience had a higher achievement on the posttest, showed better self-discipline, participation and motivation, made greater use of the multimedia, and had a better understanding of teaching goals and professional and English content, so that they were more competent in the ESP course using the multimedia courseware. In addition, such a learner-centered instruction with courseware integration was as good as that with the teacher-centered one and can offer a potential solution to overcome current problems in the development of ESP instruction in Taiwan.

Keywords
Improving classroom instruction, Interactive learning environments, Courseware integration, English for specific purposes, Higher technical education

Introduction
Given the new trend of globalization and the internationalization of the workforce, one of the goals of foreign language education must be to provide students with the foreign language ability and advanced professional knowledge necessary to succeed in the job market. This is a mandate identified by the Ministry of Education of Taiwan for technical and vocational education. ESP (English for specific purposes) instruction has accordingly become increasingly emphasized since 2000 at technical universities in Taiwan, the goal of which is to meet the needs of learners who learn English for use in their specific fields, such as business, science, technology, medicine, leisure, and academic learning (Hutchinson & Waters, 1987; Johns & Dudley-Evans, 1991).

In business, the semiconductor industry has become one of the most important industries world-wide, and over the last ten years has been offering many job opportunities in Taiwan. Thus, it is important to upgrade the level of knowledge regarding the industry’s development and simultaneously improve English skills within the current system of higher education, because the combination will help students gain related abilities, including language skills, for potential future jobs. Many in-service programs in higher education have been established in Taiwan through which adult learners can either get more job-oriented knowledge and skills or achieve self-expectations in learning more (Hsia, 2004). Such a demand provides an opportunity for the development of ESP instruction which is considered to be a learner-centered, content-based approach to teaching/learning EFL (English as a foreign language). However, there is a curious absence of discussion about teaching EFL to adults (Chang, 2004), not to mention adults’ learning behaviors and attitudes toward ESP. In addition, there are some problems in the development of ESP courses in Taiwan. After investigating the relationship of the English proficiency level of about 350 students in four universities of technology, their needs when taking ESP courses, and their expectations of an ESP teacher, Lai (2005) found that: (1) learners’ main reasons for taking ESP courses are their relevance for future jobs in business or technology, and when these students became less motivated, it was due to ineffective teachers’ conducting the subjects; (2) sufficient qualified teachers, authentic materials and specific knowledge were not provided; (3) the target need of students taking ESP courses is to be able to apply language skills such as listening,
speaking, reading and writing. Meanwhile, a recent study by Wu and Badger (2009) on analyzing teachers’ practices
in the classrooms of maritime English in a Chinese college found that what they call ISKD (In-class Subject
Knowledge Dilemma) situations happened when ESP teachers had to deal with subject knowledge with which they
are not completely familiar. The subject knowledge being delivered by an effective ESP teacher is an important issue
in understanding students’ motivation for taking ESP courses.

These above needs can be met to some extent by CALL (computer-assisted language learning) methodologies and
materials which rely on the use of interactive multimedia to integrates language skills (listening, speaking, reading,
and writing), provide authentic learning experiments, offer learners control over their learning and also focus on the
content (Ma, 2007; Tsai, 2010; Warschauer, 1996). Although courseware development and its application in
classroom lectures is becoming more greatly emphasized, its design and use have been more focused on courses
related to sciences and technology (Azemi, 2008; Jiménez & Casado, 2004; Shamsudin & Nesi, 2006). That is
because instructors in these fields have more competent skills and knowledge of multimedia software and
programming so that they are less hesitant to convert their lecture notes into an interactive package that can be
available to students. Consequently, the effectiveness of these new instructional tools has not been fully realized or
studied in ESP which is an interdisciplinary task that emphasizes coordination and integration of learning technical
knowledge and English skills. Recently, the effectiveness of CALL tools has been studied in ESP for semiconductor
technology in the four-year day-time program (DP) of the AFLD (applied foreign languages department) in a
technical University in Taiwan. The initial evaluations found that college students’ performance in courses
emphasizing courseware was as good as that under teacher-centered instruction (Tsai, 2009).

Purpose of the study

The present study was conducted to understand the effectiveness of ESP courseware integration into instruction for
adult learners of two AFLD in-service programs, a four-year night-time program (NP) and a two-year weekend
program (WP). The performance, motivation and attitude of students in these two programs were compared to teens
or young adults enrolled in DP program. In EFL environments, working with less prepared students needs to focus
on techniques that will pique student interest, stimulate a desire for additional learning experiences, and heighten
motivation. CALL tools offer that capability. In addition, in order to probe the learning effectiveness through ESP
instruction with courseware integration, a traditional teacher-centered instruction was conducted for a further
comparison.

Methods

The methodology of this study was divided into two phases, *Description of the courseware structure*, and *Integration
of courseware into instruction*, and discussed in that order.

Description of the courseware structure

The structure of the self-developed ESP courseware includes three sections (Tsai, 2009): (1) an overview, in which
three introductory topics (*Introduction to Semiconductor, Development and Application of Integrated Circuits*) are
included; (2) the core, for which seventeen topics with multimedia movies and bilingual texts (English and Chinese)
have been designed; and (3) an on-line evaluation system including listening tests and simple questions in
multimedia game-like format, combined with an instant self-checking feedback, helps learners test themselves and
monitor their learning progress and achievement. Questions have been devised for all topics related to the theories
and manufacturing technologies.

Seventeen multimedia movies with their texts are embedded in the courseware to briefly describe related theories
and manufacturing processes of the semiconductor technology mainly including atomic structure, introduction to
silicon, lattice structure, energy band structure, intrinsic and extrinsic semiconductors, preparation of wafers, thermal
oxidation, ion implantation, photolithography, etching, diffusion, and formation of thin film, bipolar diode, solar
cells, light emitting diodes and metal oxide semiconductors. An example of a session layout is shown in Figure 1.
The English text for each movie was made as brief and as accessible as possible, about 100 words for each movie, in
order to decrease text complexity and its Chinese translation was offered to allow learners a better comprehension. The terminology of the text is highlighted to reduce cognitive load, corresponding to Mayer’s signaling principle (Mayer, 2005); the color of the paragraph will change when being spoken, like karaoke style to facilitate learner’s pronunciation improvement, reading focus, and listening skills. Such a subtitled-multimedia courseware with L1 (first language: English) audio is similar to subtitled-video, which can positively promote a more efficient comprehension for L2 (second language) learners (Herron et al., 1995). Meanwhile, learners can record his or her voice through the recording program provided with Microsoft Office to practice their speaking skill.

Integration of courseware into instruction

The courseware was implemented as a seven-week module incorporated in a semester-long optional course, “English for Technology”, offered for AFLD students of three different programs in a technical university in Taiwan for four successive academic years. For the instruction with courseware integration, the course was conducted in the multimedia laboratory, and each student was assigned individually to a computer in order that he/she could study by themselves with the courseware. This study adopted a CAI (computer-aided instruction) approach combined with SCLT (sustained-content language teaching) approach where students learn language through the medium of a single content area (Murphy & Stoller, 2001). Two major components are included in SCLT: the first component concerns instruction in a specific subject area for a period of time; the second one is related to instruction in language learning, such as development of English skills, and use of learning strategies. The instruction in the first three years was conducted in a teacher-centered way without courseware integration. A learner-centered instruction for the fourth year was developed by the same teacher, by combining ESP courseware integration with teacher intervention. The curriculum design of the course was as follows:

Goal: the aim of the course was to promote students’ English vocabulary and knowledge about semiconductor technology in order to help learners improve their language and professional competence to function within the target community. Meanwhile, based upon data from pre- and post-tests, and two questionnaires, the learning effectiveness and attitude through the instruction with courseware integration was studied.
Subject: the course was taken by AFLD students of the three different programs (NP, WP, and DP). They had studied English for eight years at least: six years in junior and senior high school and two years in college. Their background and job experience were not relate to what was taught in this study. Their background is shown in Table 1.

Teacher: the teacher was the researcher who developed the courseware for the project. The teacher-researcher has a semiconductor technology academic background and has been employed in that field. For the instruction with courseware integration, the teacher played a role of supervising and observing students’ behaviors and learning, controlling the schedule, and encouraging students’ interaction with the courseware. Based on requests by individual students, the teacher also acted as a coach by giving further explanations, one-on-one, during class.

Teaching content and instruction: For the teacher-centered instruction, the teacher used a textbook (Tsai, 2011) with PowerPoint files to reinforce technological and theoretical aspects of the content that were difficult for students to understand. The courseware as a silent partner through which students took an active role to explore content knowledge and practice related linguistic fluency. In that sense, the courseware was a major medium for delivering and transferring subject content and language practices. The courseware was installed on the server in the laboratory so that students could easily access and learn the target content on their computer. Meanwhile, the teacher also used multimedia 3D visuals built into the courseware to display and explain the subject content. In general, two or three topics were given each week depending on the complexity of the topic. Since the content-specific vocabulary is important to support students’ content learning, performance of academic tasks, and use of learning strategies (Donley & Reppen, 2001), in the beginning of learning each topic for both instructions with/without courseware integration, students had to learn the technical vocabulary or terminology from the teacher’s preliminary oral explanation, and then studied the text of the topic.

Assessment: a pre- and post-test was conducted to provide students’ learning performance or evidence. The questions of these two tests were identical, but rearranged in a different order: ten simple questions asking students to briefly explain or describe terminology or process technologies, such as a p-n diode, photolithography and its process, or package and its purpose. Meanwhile, the pre-test allowed learners not only to preview the task objective, but to think ahead how to do the task and plan the knowledge and language they need which helped stimulate students’ engagement.

Questionnaire survey: After seven weeks’ instruction with the courseware integration, an internal 5-point Likert-style questionnaire of satisfaction with 14 items was administered to elicit students’ responses concerning the suitability of the courseware content and its usage, their perception of how it might support English learning and language acquisition, and their opinions about multimedia-assisted learning, and navigation of the courseware (Tsai, 2009; 2010). A second, external 5-point Likert-style questionnaire with 11 items was administered for both instructions by the academic office of the school in the end of the course to elicit information about the self-discipline and motivation of students, their self-evaluation on learning effectiveness and on teaching methodologies and materials that were delivered in class. Since these two questionnaires were not administered at the same time, students’ numbers for each questionnaire were different.

<table>
<thead>
<tr>
<th>Program</th>
<th>Student number (N)</th>
<th>Mean age</th>
<th>Mean working years</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP (Two-year Weekend Program)</td>
<td>35</td>
<td>36</td>
<td>12.3</td>
</tr>
<tr>
<td>NP (Four-year Night program)</td>
<td>30</td>
<td>24</td>
<td>1.6</td>
</tr>
<tr>
<td>DP (Four-year Day Program)</td>
<td>64</td>
<td>21</td>
<td>0</td>
</tr>
</tbody>
</table>

**Results**

**Results of students’ learning**

It was a challenging experiment and an interdisciplinary task to teach a new, practical and technical course in an AFLD department since AFLD students typically have less interest and background in science and technology. According to Gardner and Lambert (1972), attitudes and motivation have strong relation to language achievement no
matter how the learners’ aptitude and intelligence may be. Thus, how to promote students’ motivation in learning such a technical and professional course remains one of the very important concerns.

Although most students in three programs were not able to answer the questions in the pretest, they made significant progress in the posttest. The means of the students in the posttest are shown in Table 2. It indicated that the effectiveness of content learning was significantly improved after seven-week’s participation and learning for both instructions, and meanwhile the mean of the WP students in the posttest was better than those of the NP and DP students. An independent samples t-test analysis revealed that no significant difference ($p > .05$) existed in students’ posttest scores between the both instructions with/without courseware integration for each program, suggesting that students’ learning by using the student-centered instruction with courseware integration was equal to or as good as that with the teacher-centered one.

<table>
<thead>
<tr>
<th>Program</th>
<th>Instruction</th>
<th>students number (N)</th>
<th>Mean of the posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP</td>
<td>WCI</td>
<td>35</td>
<td>81.8</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>42</td>
<td>81.4</td>
</tr>
<tr>
<td>NP</td>
<td>WCI</td>
<td>30</td>
<td>73.5</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>66</td>
<td>74.9</td>
</tr>
<tr>
<td>DP</td>
<td>WCI</td>
<td>64</td>
<td>76.0</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>180</td>
<td>73.2</td>
</tr>
</tbody>
</table>

Results of students’ questionnaire of satisfaction with courseware integration

After seven-week’s learning with courseware integration, students responded to each item using a 5-point Likert scale ranging from 1 to 5, including Very Satisfied (5), Satisfied (4), Average (3), Not Satisfied (2), Disliked (1). All valid responses were input and filed for statistical data analysis using one-way ANOVA test that focused on the comparison among the three programs. An acceptable significant level for each statistics was at .05. The choices students selected for each question of the questionnaires were averaged and the standard deviation (STD) was analyzed. The Cronbach alpha reliability for the questionnaire was 0.918, indicating that the collected data were highly reliable. The results are listed in Table 3. As the satisfaction questionnaire administered by users can be considered to represent their learning motivation and effectiveness (Long, 1985; Tough, 1982), some important issues are highlighted from the means of each of the three programs:

1. The mean of consensus was 4.01, indicating that the majority of the students in the three programs are satisfied with the integration of the ESP courseware into instruction. Among all the means of the three programs, there were 9 questions higher than or equal to 4.00 which appeared that students hold the affirmative opinions to these questions.

2. QF1 (teaching with the courseware, $M = 4.20$) had the highest score. This result meant that the instruction with courseware integration was supported by the students of all three programs. In fact, the fullest collaboration for ESP teaching is often said to be one where a subject expert and a language teacher team-teach classes (Johns & Dudley-Evans, 1991). However, such teaming has not been feasible in vocational education in Taiwan for several reasons, such as lack of qualified teachers, difficulties of collaboration or relevant curriculum design. The ESP courseware incorporating L1 audio with paragraph subtitles can be considered as an ‘assistant’ ESP teacher which not only helps Chinese students of English practice language skills such as pronunciation, spelling, listening, reading, translation and short question writing, but also helps promote students’ professional comprehension through their interaction with its multimedia content.

3. The highest score also for QF3.7 (terminology learning, $M = 4.20$) revealed that, increasing vocabulary comprehension is seen as the most effective and important reading strategy which allows students to have a better understanding of lectures and texts (DeCarrico, 2001), and learning of the content-specific vocabulary should be emphasized in the ESP courseware design and during its integration into instruction, which has semantic ties and conceptual relationships with the target content.

4. QF6 (multimedia assisted learning, $M = 4.20$) also had the highest score. It indicated that the inclusion of multimedia leads to enhance students’ learning, and the layout of the multimedia movies with their brief and
accessible texts played a scaffolding role to help students develop or improve their professional ideas or knowledge that was reinforced by a higher score in QF2 (improvement in professional knowledge, M = 4.02).

5. QF8 (navigation, M = 4.12) had a higher score, suggesting that the user-friendly learning interface design provided guidance into multimedia environment that allowed students to learn leads (Mayer, 2005). Moreover, such a design with multimedia inclusion can not only ease off learning pressure and increase cognitive ability, but also add pedagogical value to the application and promote learner’s motivation which has been considered one of the key factors in second language learning (Oxford and Nyikos, 1989).

6. A higher score for QF3.6 (translation, M = 4.09) indicated that translation can help Chinese students of English get a better comprehension of professional knowledge in studying such a technical ESP course, further enhanced by corresponding multimedia movies (QF6) and the teacher’s intervention (QF1). Besides, translation of technical terms and texts provided by the courseware can reduce or avoid unnecessary mistakes and misunderstanding from word by word translation.

7. The scores for learning English skills from QF3.1 to QF3.5 are 3.68 - 4.01, higher than the average, or 3. It indicated that the courseware provided enough practices for training integrative English skills, which can meet the need of Taiwanese EFL students in learning this ESP courses (Lai, 2005).

<table>
<thead>
<tr>
<th>Questions</th>
<th>WP (N=24)</th>
<th>NP (N=22)</th>
<th>DP (N=48)</th>
<th>Mean of the three programs</th>
<th>One-way ANOVA among three programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>QF1. You are satisfied with the teaching with the courseware.</td>
<td>4.58 (.615)</td>
<td>4.05 (.486)</td>
<td>4.08 (.647)</td>
<td>4.20</td>
<td>7.029 WP&gt;NP** WP&gt;DP**</td>
</tr>
<tr>
<td>QF2. The courseware improves your professional knowledge in semiconductor?</td>
<td>4.17 (.565)</td>
<td>3.95 (.653)</td>
<td>4.00 (.619)</td>
<td>4.03</td>
<td>.818</td>
</tr>
<tr>
<td>QF3. Overall</td>
<td>4.42 (.584)</td>
<td>3.77 (.685)</td>
<td>3.92 (.539)</td>
<td>4.01</td>
<td>8.161 WP&gt;NP** WP&gt;DP**</td>
</tr>
<tr>
<td>QF3.2. Listening Skills?</td>
<td>4.08 (.562)</td>
<td>4.00 (.617)</td>
<td>3.69 (.719)</td>
<td>3.86</td>
<td>3.464 WP&gt;DP*</td>
</tr>
<tr>
<td>QF3.3 Speaking Skills?</td>
<td>4.04 (.624)</td>
<td>3.55 (.800)</td>
<td>3.56 (.741)</td>
<td>3.68</td>
<td>3.963 WP&gt;DP*</td>
</tr>
<tr>
<td>QF3.4 Reading Skills?</td>
<td>4.33 (.565)</td>
<td>3.77 (.612)</td>
<td>3.92 (.710)</td>
<td>3.99</td>
<td>4.830 WP&gt;NP** WP&gt;DP*</td>
</tr>
<tr>
<td>QF3.5 Writing Skills?</td>
<td>3.92 (.504)</td>
<td>3.59 (.590)</td>
<td>3.65 (.601)</td>
<td>3.70</td>
<td>2.309</td>
</tr>
<tr>
<td>QF3.6 Translation Skills?</td>
<td>4.42 (.654)</td>
<td>3.91 (.811)</td>
<td>4.00 (.684)</td>
<td>4.09</td>
<td>3.655</td>
</tr>
<tr>
<td>QF3.7 Vocabulary of Technology?</td>
<td>4.46 (.588)</td>
<td>4.09 (.684)</td>
<td>4.13 (.672)</td>
<td>4.20</td>
<td>2.487</td>
</tr>
<tr>
<td>QF4. The professional content of the courseware is relevant.</td>
<td>4.25 (.532)</td>
<td>3.73 (.150)</td>
<td>4.00 (.619)</td>
<td>4.00</td>
<td>4.093 WP&gt;NP*</td>
</tr>
<tr>
<td>QF5. The English of the courseware is relevant.</td>
<td>4.42 (.584)</td>
<td>3.91 (.684)</td>
<td>3.85 (.583)</td>
<td>4.01</td>
<td>7.250 WP&gt;NP* WP&gt;DP*</td>
</tr>
<tr>
<td>QF6. The multimedia animated movies of the courseware assist learning.</td>
<td>4.50 (.590)</td>
<td>3.95 (.785)</td>
<td>4.17 (.753)</td>
<td>4.20</td>
<td>3.381 WP&gt;NP*</td>
</tr>
<tr>
<td>QF7. The evaluation system of the courseware assists learning.</td>
<td>4.29 (.690)</td>
<td>3.77 (.685)</td>
<td>3.90 (.722)</td>
<td>3.97</td>
<td>3.619 WP&gt;NP*</td>
</tr>
<tr>
<td>QF8. The function keys improve navigation.</td>
<td>4.25 (.608)</td>
<td>3.86 (.774)</td>
<td>4.17 (.630)</td>
<td>4.12</td>
<td>2.238</td>
</tr>
</tbody>
</table>

The overall average score 4.30 3.85 3.93 4.01

*: p<.05; **: p<.01
If compared by programs, WP students who were older and had greater working experience had higher means for all the questions in the satisfaction questionnaire than NP and DP students by 0.22-0.65. Some studies found (Hsia, 2004; Wang, 2003; Yu, 1998) that learners taking in-service programs at the age of 30-40 generally had higher instrumental motivation, which is taken as a desire to use the target languages to attain practical goals, such as getting a job or reading technical materials (Hudson, 2000), so that they showed a better learning satisfaction.

A further analysis through one-way ANOVA among the programs indicated that there was no significant difference between NP and DP, probably due to their similar background. However, a significant difference existed among WP, NP and DP in several questions related such as QF1 (teaching with the courseware), QF3.1 (overall English skills), QF3.2 (listening), QF3.3 (speaking), QF3.4 (reading), QF4 (professional content), QF5 (English content), QF6 (multimedia-assisted learning) and QF7 (on-line evaluation), shown in Table 3. Since WP had a higher posttest score than the other two, the result indicated that WP students with higher achievement were more satisfied with the instruction, and professional and English content provided by the courseware, and made more use of learning activities such as practices of English skills, multimedia and on-line evaluation. Related findings had also been observed in the earlier study: the students with high scores had a better participation, understanding and satisfaction while studying with courseware integration so that they were more competent to learn in such a professional and learner-centered ESP course with the courseware (Tsai, 2009).

Results of the external on-line questionnaire

An external on-line questionnaire with 11 items was administered, including three parts: students’ self-discipline and motivation (items 1 and 2), participation and instruction (items 3, 4, 5, 6, 7, 10 and 11) and students’ self-evaluation of learning effectiveness (items 8 and 9). A 5-point Likert scale was used and the Cronbach alpha reliability for the questionnaires was 0.951. There were respectively 78 and 271 students for the instructions with/without courseware integration who validly completed the questionnaire. The result is listed in Table 4.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Instruction</th>
<th>Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1. I pay attention in class.</td>
<td>WCI</td>
<td>4.46</td>
<td>.502</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.59</td>
<td>.522</td>
</tr>
<tr>
<td>E2. I always attend the class.</td>
<td>WCI</td>
<td>4.31</td>
<td>.492</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.41</td>
<td>.659</td>
</tr>
<tr>
<td>E3. What is taught corresponds to the teaching goal.</td>
<td>WCI</td>
<td>4.37</td>
<td>.647</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.41</td>
<td>.659</td>
</tr>
<tr>
<td>E4. The teaching materials and handouts are rich.</td>
<td>WCI</td>
<td>4.36</td>
<td>.664</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.38</td>
<td>.729</td>
</tr>
<tr>
<td>E5. The teaching content is relevance.</td>
<td>WCI</td>
<td>4.37</td>
<td>.729</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.37</td>
<td>.733</td>
</tr>
<tr>
<td>E6. The teacher makes use of the material and gives a systematical explanation.</td>
<td>WCI</td>
<td>4.38</td>
<td>.649</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.42</td>
<td>.715</td>
</tr>
<tr>
<td>E7. The teacher gives examples or uses auxiliary tools to facilitate learning.</td>
<td>WCI</td>
<td>4.40</td>
<td>.690</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.41</td>
<td>.704</td>
</tr>
<tr>
<td>E8. I get some professional knowledge and skills.</td>
<td>WCI</td>
<td>4.26</td>
<td>.763</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.38</td>
<td>.755</td>
</tr>
<tr>
<td>E9. The abilities of thinking, analysis and problem-solving are improved.</td>
<td>WCI</td>
<td>4.35</td>
<td>.735</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.41</td>
<td>.693</td>
</tr>
<tr>
<td>E10. The preparation and attitude of the teacher is serious and careful.</td>
<td>WCI</td>
<td>4.42</td>
<td>.614</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.53</td>
<td>.643</td>
</tr>
<tr>
<td>E11. I recommend others to take this course.</td>
<td>WCI</td>
<td>4.37</td>
<td>.705</td>
</tr>
<tr>
<td></td>
<td>F2F</td>
<td>4.49</td>
<td>.688</td>
</tr>
</tbody>
</table>
All the means of the questions for these two instructions were higher than 4, meaning students were satisfied with their individual instruction. Moreover, their means for all the questions were very close to each other. After a further analysis through the independent samples t-test, there was no significant difference between the two instructions. Since no significant difference was either found in students’ posttest scores between both instructions for each programs, it implied that courseware integration into ESP instruction in this study did play a role of an content and language tutor with which students performed as well as those in a teacher-centered way.

The means of the question E6, E7 and E11 related to the teacher’s teaching attitude and strategies in both instructions were higher than 4.38, meaning that the teacher-researcher with academic background and working experience about semiconductor technology had enough ability to handle both language and content, and succeeded in leading students to an effective and meaningful ESP in both instructions by giving good preparation, making use of the material and auxiliary tools and giving systematical explanation. The positive teaching attitude and belief of the teacher was considered as a key factor for success in conducting such a technical and professional ESP course.

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean (STD)</th>
<th>F test</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>WCI 4.54 (.539)</td>
<td>4.35 (5.26)</td>
<td>4.42 (.519)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.62 (.519)</td>
<td>4.59 (.514)</td>
<td>4.58 (.498)</td>
</tr>
<tr>
<td>E2</td>
<td>WCI 4.62 (.741)</td>
<td>4.18 (.747)</td>
<td>4.27 (.598)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.50 (.506)</td>
<td>4.44 (.393)</td>
<td>4.37 (.494)</td>
</tr>
<tr>
<td>E3</td>
<td>WCI 4.92 (.544)</td>
<td>4.35 (.706)</td>
<td>4.23 (.772)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.60 (.277)</td>
<td>4.47 (.493)</td>
<td>4.34 (.692)</td>
</tr>
<tr>
<td>E4.</td>
<td>WCI 4.85 (.547)</td>
<td>4.29 (.728)</td>
<td>4.25 (.801)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.57 (.376)</td>
<td>4.47 (.288)</td>
<td>4.29 (.700)</td>
</tr>
<tr>
<td>E5</td>
<td>WCI 4.92 (.550)</td>
<td>4.24 (.747)</td>
<td>4.19 (.761)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.55 (.277)</td>
<td>4.44 (.664)</td>
<td>4.29 (.790)</td>
</tr>
<tr>
<td>E6</td>
<td>WCI 4.92 (.508)</td>
<td>4.29 (.558)</td>
<td>4.27 (.785)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.71 (.277)</td>
<td>4.56 (.588)</td>
<td>4.29 (.676)</td>
</tr>
<tr>
<td>E7</td>
<td>WCI 4.92 (.526)</td>
<td>4.35 (.587)</td>
<td>4.27 (.761)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.67 (.277)</td>
<td>4.55 (.702)</td>
<td>4.29 (.707)</td>
</tr>
<tr>
<td>E8</td>
<td>WCI 4.77 (.570)</td>
<td>4.00 (.639)</td>
<td>4.21 (.813)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.67 (.439)</td>
<td>4.50 (.791)</td>
<td>4.26 (.771)</td>
</tr>
<tr>
<td>E9</td>
<td>WCI 4.77 (.517)</td>
<td>4.12 (.662)</td>
<td>4.31 (.723)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.69 (.439)</td>
<td>4.48 (1.054)</td>
<td>4.31 (.624)</td>
</tr>
<tr>
<td>E10</td>
<td>WCI 4.92 (.484)</td>
<td>4.24 (.519)</td>
<td>4.35 (.703)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.76 (.277)</td>
<td>4.62 (.562)</td>
<td>4.44 (.635)</td>
</tr>
<tr>
<td>E11</td>
<td>WCI 4.85 (.544)</td>
<td>4.18 (.523)</td>
<td>4.31 (.754)</td>
</tr>
<tr>
<td></td>
<td>F2F 4.74 (.376)</td>
<td>4.61 (.636)</td>
<td>4.37 (.748)</td>
</tr>
</tbody>
</table>

overall WCI 4.82 4.25 4.28 4.36
In addition, the means of the questions in the questionnaire for each program in both instructions are shown in Table 5, all the means of the questions higher than 4. WP students with better posttest scores had higher means for all the questions than NP and DP students in both instructions. Taking the one with courseware integration into consideration, a further one-way ANOVA analysis indicated that there was no significant difference between DP and NP where students’ posttest scores were close to each other, but a significant difference among WP, DP and NP, existing in E2 (attendance), E3 (understanding of teaching goal), E4 (teaching materials and handouts), E5 (relevant content), E6 (teacher’s instruction and explanation), E7 (use of auxiliary tools), E10 (teacher’s attitude), and Q11 (recommendation), shown in Table 5. These differences revealed that better understanding about learning goal, more students’ active participation and more supportive attitude from the teacher has a significantly positive influence on the learning satisfaction and effectiveness of WP students.

In case of instruction without courseware integration, similar significant difference was also found in E6 – E11, especially between adult WP students and college DP students. Threlkeld and Brzoska (1994), studying distance education which is another kind of learner-centered approach, noted that “maturity, high motivation levels, and self-discipline have been shown to be necessary characteristics of successful, satisfied students” (p. 53). Their findings could explain what was found for the better performance of WP students in this study.

**Discussion**

The ESP instruction in this study was respectively conducted without and with courseware integration to students of three different programs. Although students were satisfied with F2F instruction without courseware integration, some problems were observed: (1) it was difficult for a teacher to apply integrative training in English skills within a class, especially with about 60 students in a class; (2) Even though terminology, principles, and manufacturing processes were introduced, the instruction became boring and inefficient due to students’ repeated questions about subject knowledge or its basic chemical or physical principles, rather than about linguistic knowledge.

To remedy the problems mentioned above, integrating courseware incorporating L1 audio with paragraph subtitles and its Chinese translation became a possible solution. During the seven-week’s learning process, the computer was a major medium for delivering and transferring content knowledge and language practices, and the courseware played three main roles: (1) Tutor: making the transfer of professional knowledge by providing texts with L1 audio and multimedia movies; (2) Language teacher: providing learners with repeated practices of language skills; (3) Peer-like role with which students can have a direct interaction throughout their learning process. Other beneficial features have been found: First, students were able to pay attention to the input from their direct and individual interaction with the courseware within the classroom. Second, students had more freedom and responsibility to learn at their own pace and need. Meanwhile, it gave more equal participation and opportunity for students to study the target content and practice its related language skills. Third, students felt less shy to ask the teacher for one-on-one explanations, and meanwhile the teacher was more available for such individual requests.

Of course, this integration is not a panacea and has its limitations such as requiring investment of time and money, interdisciplinary collaboration and integration in courseware production. Moreover, it brings no guaranteed results. However, when appropriately implemented, integrating courseware into instruction can offer the means to conduct a more thorough integration of language and content knowledge, and provide students with unprecedented opportunities for autonomous learning.

According to the questionnaire survey, students’ satisfaction with practices of integrative English skill reinforced the layout and design of the courseware in language. The integrated ESP courseware in this study met Chapelle’s suggestions (1998) for multimedia CALL based on hypotheses about ideal conditions for SLA (second language acquisition). For example, the color change of the paragraph and its audio and Chinese translation can be conducted and provided by just clicking the right or left buttons of the mouse. These features correspond to Chapelle’s first suggestion mentioning making key linguistic characteristics salient by highlighting them in a different color, in aural input, or transcription of phrases containing linguistic elements.
Chapelle’s second suggestion concerns linguistic input provided through either written or aural language and modified by several forms such as repetition, simplification through restatements, non-verbal cues, decreased speed, reference materials, and change of input mode. This suggestion can be achieved by repeatedly practicing integrative skills with written English texts combined with L1 audio and related multimedia movies offered by the courseware. These types of linguistic modifications are distinct from the materials found on the Web because they hold the potential to provide learners with comprehensible input rather than just input. In addition, the language training offered by the on-line evaluation system of the courseware provide learners with opportunities for comprehensible output which can be conducted either written or spoken by using target language forms to stretch their competence, as mentioned in Chapelle’s third suggestion. Besides, the instant self-check function of the online evaluation system allows learners to analyze, recheck, reflect, and identify, and even correct their errors. This design corresponds to Chapelle’s fourth and fifth suggestions regarding the provision of opportunities for learners to notice their errors and correct their linguistic output.

The sixth and the last suggestions imply supporting modified interaction between the learner and the computer and acting as a participant in L2 tasks. It can be accomplished by several functions provided by the courseware through mouse clicks, hypertext links, and a variety of learning activities in language and subject content. Students’ learning effectiveness was presented in the results. While the courseware was integrated into instruction, the task assigned in this study focused the learner's achievement on accomplishing a Q&A test through the use of language rather than only on solving problems of linguistic form, as suggested by Chapelle.

Besides, no matter the instruction with or without courseware integration, students had less confidence to express themselves in English. Moreover, they expected that the teacher could explain complicated and unclear parts in Chinese for better understanding. Accordingly, bilingual explanations for difficult or abstruse technical knowledge in English and Chinese was sometimes useful and necessary because teacher’s explanation with familiar Chinese technical terms that they had learned before could reduce students’ cognitive load and anxiety in subject knowledge and enhance their comprehension in studying such a technological ESP course. Moreover, the teacher with experience of teaching English, academic background and working experience had enough ability to handle both language and content, and succeeded in leading students to an effective meaningful learning. Lai’s (2005) survey agree on the importance of having effective ESP teachers who are able to connect to learners with real work experience, provide knowledge related to jobs, exercise English skills, and offer authentic and meaningful materials.

**Conclusion**

Expanded abilities and more positive attitudes toward e-learning are important new literacy that most educational institutions now urge students to establish. Integrating courseware into instruction within the classroom shifts the teacher-centered learning to student-centered learning and this approach focuses on helping students become autonomous learners who can continue to learn how to communicate, conduct research or present ideas effectively in their life-long or continuing learning after graduating from school. Such students’ learning ownership is especially important in an interactive multimedia environment where students have to engage more actively in the cognitive processes of selecting, organizing, integrating and applying what they acquire in the learning process. Meanwhile, the teacher offering instruction with courseware integration needs to play a role of facilitating students’ learning process, encouraging and enabling them to study with the courseware to be independent and responsible learners.

The main objective of the ESP course in this study is to help students establish basic knowledge or understanding of semiconductor technology, more emphasizing terminology and content learning. In fact, it is a tough task for ESP teachers to meet and conduct all these requirements within the classroom, especially in teaching more technical ESP courses. Thus, the utilization of instructional tools should be seriously considered, depending on the nature and goals of the courses, to conduct effective and smooth instruction. Based on the constructivist approach, ESP instruction with courseware integration in this study allowed students to interact with the content, to explore and construct vocabulary and meanings. Such individualization in learning enhanced by technology has been proved in this study. Of course, more classroom-oriented research is required to determine the full impact of such courseware integration and gain more insights.

According to the results of the posttest and students’ responses to the internal questionnaire of satisfaction and the external on-line questionnaire of self-evaluation, the performance, motivation and attitude of adult students have
been analyzed and were compared with those of college ones. Meanwhile, the learning effectiveness through the instruction with ESP courseware integration was also studied. In conclusion, some important findings include these:

1. Like college students, adult students of the in-service programs were satisfied with the courseware which provided professional knowledge with practices of English skills including vocabulary usage, pronunciation practice, speaking opportunities, reading and translation practice, listening practice and short answer writing exercises. It suggested that the courseware played the role of an adjunct content and language teacher.

2. Professional knowledge and English skills of students in all three programs were improved from the learner-centered instruction with courseware integration. Moreover, students’ performance in the posttest was as good as that in the teacher-centered instruction.

3. Multimedia movies embedded in the courseware supported a friendly multimedia-assisted environment which acted out real-life experience so that the transfer of information and knowledge to learners can be facilitated and promoted both visually and auditorily. Furthermore, such multimedia inclusion combined with user-friendly navigation design enhanced knowledge transfer for students and promoted learner’s motivation which led to a better learning satisfaction and effectiveness.

4. Students in the WP program who were older and had greater working experience performed better in the posttest than those of the NP and DP programs in both types of instruction with/without courseware integration. The questionnaire results indicated that WP students showed better self-discipline, participation and motivation. Meanwhile, they made more use of multimedia and had a better understanding of teaching goal and professional and English content so that they were more competent to learn such a learner-centered technological ESP course with courseware integration.

5. The team instruction combining well-designed ESP courseware and subject teacher’s intervention did offer a potential solution to problems in the development of ESP courses in Taiwan to meet learners’ needs in professional knowledge and integrative language skills. In case fullest collaboration for ESP teaching is not feasible, more technical ESP courses can be taught by subject teachers with experience of teaching English and real work so that ISKD situations would be reduced or avoided and learners’ effective and meaningful learning in the desired specific field or purpose can be conducted better and more smoothly.

Acknowledgements

The author would like to thank Mr. Lin Jinde for his help in the multimedia design and production of 3D animated movies. Meanwhile, the author would like to thank most sincerely Dr. Boyd Davis, Professor of Applied Linguistics/English, University of North Carolina-Charlotte, for her valuable suggestions.

References


An Empathic Avatar in a Computer-Aided Learning Program to Encourage and Persuade Learners

Gwo-Dong Chen, Jih-Hsien Lee, Chin-Yeh Wang*, Po-Yao Chao¹, Liang-Yi Li and Tzung-Yi Lee

Department of Computer Science & Information Engineering, National Central University, Taiwan // ¹Department of Information Communication, Yuan Ze University, Taiwan // chen@csie.ncu.edu.tw // jhlee@iii.org.tw // chinyea@db.csie.ncu.edu.tw // lihenry12345@gmail.com // bear@db.csie.ncu.edu.tw // poyaochao@saturn.yzu.edu.tw

*Corresponding author

(Submitted July 23, 2009; Revised January 26, 2011; Accepted March 31, 2011)

ABSTRACT

Animated pedagogical agents with characteristics such as facial expressions, gestures, and human emotions, under an interactive user interface are attractive to students and have high potential to promote students’ learning. This study proposes a convenient method to add an embodied empathic avatar into a computer-aided learning program; learners express their emotions by mouse-clicking while reading, and the avatar motivates them accordingly. This study designs empathic responses for avatars to encourage and persuade learners to make greater reading effort. This experiment examines emotional recognition, empathy transformation, and the effect of virtual human encouragement and persuasion. Subjects identify facial expressions of the avatar, especially those expressing positive facial emotions. Compared to the contrast group, the empathic avatar increases learners’ willingness to continue reading and complete exercises.

Keywords

Animated pedagogical agent, Computer-aided learning, Encourage, Persuade, Empathy

Introduction

Animated pedagogical agents (APAs), with characteristics such as facial expressions, gestures, human emotions, and an interactive user interface, are attractive to students. Many studies posit that social agencies with social cues in multimedia messages encourage learners to interpret human–computer interactions as a parallel to human-to-human conversation. These agents provide students with more lifelike interactions that could increase the communication capacity of learning systems and the ability of these systems to engage and motivate students. Many learning environments have been integrating APAs since their conception to encourage and motivate students to make greater learning effort.

Lester, Converse, Kahler, Barlow, Stone, & Bhogal (1997a) and Lester, Converse, Stone, Kahler, & Barlow (1997b) developed an interactive learning system with a lifelike agent in his laboratory by a large, multidisciplinary team of computer scientists, graphic designers, and animators. The lifelike agent provides students with customized advice in response to their problem-solving activities and plays a critical motivational role as it interacts with students by giving principle-based animated advice to challenge the student, or employs task-specific advice if the student is having difficulty. Experimental results revealed that students perceive the agents as being helpful, credible, and entertaining. Emotions in the agent design reinforce its expression of social cues and satisfy learners’ emotional needs. Picard (1997), in the book Affective Computing, indicates that emotions play an essential role in decision making, perception, learning, and more—they influence the very mechanisms of rational thinking. Lester, Towns, & FitzGerald (1999) used lifelike pedagogical agents with visual emotive communication (including facial expression, full-body behaviors, arm movements, and hand gestures that visually augment verbal problem-solving advice and encouragement) to encourage and motivate students. D’Mello et al. (2008) revealed an affect-sensitive Intelligent Tutoring System, which detects learner emotions by monitoring conversational cues, gross body language, and facial features by hardware. Their research considered learners’ affective and cognitive states in selecting pedagogical and motivational dialogue moves, and responding through an embodied pedagogical agent with animated facial expressions and modulated speech. McQuiggan, Rowe, Lee, & Lester (2008) showed a narrative-centered learning
environment and evaluated its affective transitions. Jaques, Lehmann, & Pesty (2009) created an emotional pedagogical agent with affective tactics that predicts students’ emotions according to the psychological model OCC (see Ortony & Collins, 1988). Though the prediction rate of specific emotions can reach a certain high percentage, improper emotional responses based on the fault of automatic emotion prediction hypothesis might provoke negative feelings in users.

Many studies have mixed multimedia materials with APAs for increasing students’ motivation and learning effect. System development requires a large, multidisciplinary team of domain experts, computer scientists, graphic designers, and animators (Lester et al., 1997a). Johnson, Rickel, & Lester (2000) listed eight types of human-APA interactions that could benefit learning environments, including interactive demonstrations, navigational guidance, gaze and gesture as attentive guides, nonverbal feedback, conversational signals, conveying and eliciting emotion, virtual teammates, and adaptive pedagogical interactions. Many researchers have created animated pedagogical agents who convey emotional responses to the tutorial situation to increase the learning environment by engaging and motivating students. These studies necessitate dedicated and complicated detectors and calculating formulas to work together with learning systems for active and precise detection of users’ emotions by facial expressions, full-body behaviors, arm movements, or hand gestures (D’Mello et al., 2008). These requirements hinder the materialization in general subjects and practice. To reduce the complexity of detecting users’ emotions, some systems inferred their emotions according to the psychological model (Jaques, Lehmann, & Pesty, 2009).

To present APAs, some research has adopted 2D or 3D animation approaches. Okonkwo & Vassileva (2001) designed an emotional pedagogical agent with simplified facial expressions that motivates students by convincing them it really cares about their performance and progress throughout the training. Kim, Baylor, & Shen (2007) created portrait-sized pedagogical agents as learning companions, developed using a 3D-animation-design tool to provide context-specific information and helpful messages at learners’ request. Experimental results showed that APA’s empathetic responses had a positive impact on learner interest and self-efficacy. The social agency delivering the verbal and visual social cues, such as facial expressions, gestures, head and body movements of the animated agent, in computer-based environments, fosters the development of a partnership by encouraging learners to consider their interaction with the computer to be similar to that of a human-to-human conversation (Mayer, Sobko, & Mautone, 2003; Moreno, Mayer, Spires, & Lester, 2001). Although many studies have investigated APA approaches, it is still worth investigating in how APAs affect students’ emotions and behaviors in different learning activities.

This research evaluates how emotional avatars in different learning activities affect students’ e-learning behavior. We added an embodied empathic avatar into a computer-aided learning program; learners express their emotions by mouse clicking while learning, and an avatar cares about and motivates them accordingly by its coherent voice and upper-body motions. Emotional buttons set beside the e-learning system screen passively read users’ emotions after they press the emotional buttons. Empathic responses from avatars encourage and persuade learners to increase effort. The results of this study may support existing pedagogical theory and help utilize technique strategies in promoting an e-learning system with APAs.

The organization of the paper is as follows: Section 2 describes the empathic avatar design and learning system; Section 3 shows experiments and results; Section 4 and 5 present discussion and conclusions.

**Empathic avatar design**

Avatars should be polite (Wang, Johnson, Mayer, Rizzo, Shaw, & Collins, 2008) and emotionally positive (Kim, Baylor, & Shen, 2007), possess a real voice with full social cues (Atkinson, Mayer, & Merrill, 2005), and demonstrate proper hand gestures for indications that might impact on learners (Baylor & Kim, 2009). This research considers trust relationship between agents and learners as one of the factors influencing motivation (Fogg, 2003). An emotional-interactive system design should also consider instructions. This study focuses on how to add an empathic avatar to an existing learning system to promote students’ learning, and uses textbooks in the system as learning materials. The avatar plays the role of a learning companion who encourages and persuades students to make more effort in learning. The user interface of the system allocates the main area on the computer screen for students’ learning and reserves the right column for human-computer interaction, including emotional expressions and animated tutor reactions. Figure 1 shows the system’s user interface. Students read and do exercises on the left side. They can also switch learning activities anytime. Predicting the emotions of students from various backgrounds
and diverse cognitional and emotional statuses is difficult. A blank area allows learners to express their emotions at any time, and an animated avatar appearing in the video display area responds to assist in their learning (see the upper-right area in Fig. 1). Learners express their emotions by clicking one of the four-pair emotional buttons (see the bottom-right area in Fig. 1). Four pairs of basic negative-positive emotions occurring during interaction with the educational system (Tzvetanova & Tang, 2005) include the emotions of distress-joy, fear-confidence, boredom-fascination, and unsatisfied-satisfied. Choosing the emotion set simplifies students’ choice of emotional expressions and covers most emotions while learning. The middle-right area of Figure 1 is the record of the student-avatar dialog.

Figure 1. Interface of the system

The following three sessions describe how to create the virtual human and how the virtual human interacts with students.

Creating the avatar

According to the social agency experiment, APAs might imitate human interactions as much as possible to enhance the sensation by delivering large numbers of social cues (Mayer et al., 2003; Moreno et al., 2001). Nguyen & Canny (2009) have also shown that upper-body video framing, which conveys subtle cues, might be more effective than face-to-face interaction regarding empathy effects. However, mimicking real humans requires that professionals spend much time on designing a vivid, detailed virtual human. This research chose 3D animation software to create the virtual human and its performance, while considering the appeal of 3D roles with various built-in emotional expressions. Development of the virtual-human animation focuses on behaviors and voices. A class of 52 college students elected built-in 3D roles as their favorite, and selected a female-embodied virtual human named Maggie. Poser 6, a 3D animator, was used to create the actor-empathic performance. After the animations were created, they were converted to MPEG4 (AVI-Format) films for integrating into the learning system.

Figure 2. Examples of the avatar’s facial emotions
Adding empathic characteristics to the virtual human involves adding facial expressions, hand gestures, body movements, and voices. Ekman & Friesen (1978) interpreted that each emotion puts unique signals in the face, so facial expressions are more reliable indicators of a person’s emotional state than is body language. This study adopted Ekman’s facial action coding system in the design. The basic six emotions include happiness, anger, surprise, sadness, disgust, and fear. To feed the needs of e-learning, two emotions (worry and neutral) were also included. Figure 2 shows five of the avatar’s facial emotions.

This work referred to studies on human affective gestures (Givens, 2002) for hand gestures and body movements to make the avatar’s nonverbal behaviors consistent with its empathic expressions. Figure 3 shows two poses of the design. This project recorded the avatar voices from a dub actor.

**Interaction design of the avatar**

The purpose of the student-avatar interaction is to encourage and persuade users to increase engagement and learning effect. Greetings and self-introduction in the beginning serve as orientation; caring, and empathic responses are designed in the passive mode, which are awakened after students trigger the emotional expression or after an examination; and a farewell greeting and well wishes end the session.

The avatar is designed as a friend to listen to students, to care for and encourage them. Parallel-empathic responding strategy is used for caring and encouragement. The designed conversation scenarios are as listed below.

Greetings: When a student logs into the system, the avatar appears in the interaction area and welcomes the user with a smile.

Empathic responses: Emotional buttons beside the e-learning screen obtain users’ emotions after they press the emotional buttons. The avatar always wears a smile. After students express their emotions by pushing one of the buttons, the avatar uses empathy to encourage and persuade them into persistent learning. Besides these empathic reactions, the avatar persuades students to keep learning when they want to break from learning activities. The next session describes the details.

Farewell: When students want to leave, the avatar waves and says, “Goodbye, take care, and hope we can meet again.”

<table>
<thead>
<tr>
<th><strong>Table 1. Empathic responses of the avatar</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative emotions</strong></td>
</tr>
<tr>
<td><strong>(Status: Distress)</strong></td>
</tr>
<tr>
<td>Voice: I feel sad when I hear that you are distressed.</td>
</tr>
<tr>
<td>Cheer up, never give up.</td>
</tr>
<tr>
<td>Facial emotion: from Sadness to Neutral</td>
</tr>
<tr>
<td>Voice: I feel sad to see that you have fear. But don’t worry too much. Remember to keep learning.</td>
</tr>
<tr>
<td>Facial emotion: from Sadness to Fear, and then Neutral</td>
</tr>
<tr>
<td><strong>(Status: Fear)</strong></td>
</tr>
<tr>
<td>Voice: Do the learning activities make you feel bored? Sometimes we need to persist to the end to gain knowledge.</td>
</tr>
<tr>
<td>Facial emotion: from Worry to Neutral</td>
</tr>
<tr>
<td><strong>(Status: Boredom)</strong></td>
</tr>
<tr>
<td>Voice: I am sad to see you dissatisfied. No matter what has happened, never mind.</td>
</tr>
<tr>
<td>Facial emotion: from Sadness to Neutral</td>
</tr>
</tbody>
</table>
Encourage and persuade students with empathy into persistent learning

To design avatars to encourage and persuade students to continue learning, maintaining the relationship between avatars and students is necessary. Besides greetings, encouragement, and saying goodbye, we considered empathic reactions in designing avatars. “Empathy is the process of putting oneself in the place of another person, seeing matters from the other’s perspective, perceiving the other’s feelings and thoughts, and conveying this awareness to that person” (Davis, 1996). McQuiggan & Lester (2007) also described the processes of empathy as follows. First, the antecedent consists of the empathizer’s consideration of herself, the target’s intent and affective state, and the situation at hand. Second, assessment consists of evaluating the antecedent. Third, empathic outcomes, for example, behaviors express concern. Table 1 shows one set of avatar empathic responses when students signal their emotions.

For example, when a student tells the system that he is distressed, the avatar responds, “I feel sad when I hear that you are distressed,” with a sad face and two arms crossed in front of her chest. The avatar continues saying, “Cheer up. Never give up,” in a confirming intonation accompanied by a neutral facial expression, arms bent at the elbow and raised, and hands in a fist (see the left part and the right part in Figure 3).

![Figure 3. Nonverbal expressions of the avatar](image)

Besides these empathic reactions, the avatar persuades students to keep learning when they want to break from learning activities. When students indicate that they want to stop reading by pressing the upper-middle button labeled Stop Learning, the avatar will say, “You are great. You have learned so much, but do you really want to leave? Don’t you want to learn more?” When students want to quit doing exercises, the avatar says, “You are great. You have done so many exercises, but do you really want to quit? Don’t you want to do more exercises?”

Experiment

To examine the avatar effect, this study first examined whether students could recognize the avatar’s facial emotion in Experiment 1. In Experiment 2, we examined whether students could tell how the avatar felt, based on its upper-body performance, and whether or not students could feel empathy from the virtual human. This work also examined whether encouragement and persuasion help students put more effort into learning activities.

Experiment 1

The purpose of the experiment was to determine whether users could recognize the emotions behind the agent’s facial expressions. The results of this experiment designed to elicit feedback from users were used to improve the quality of our emotional agent design.

The subjects were 52 college students from the department of computer sciences at National Central University. They were freshmen and had been using computers for five years on average, and only a few of them had ever interacted with avatars. The eight facial expressions of the avatar were shown separately, and the subjects were asked to rate their perception level of the eight emotions expressed on five-point scales. For example, the first question assessed the user’s perception level: “Do you agree that the facial emotion of the virtual human is anger?” Choices

Choices
ranged from 1 (strongly disagree) to 5 (strongly agree). Besides rating emotions, subjects were encouraged to give their opinions about the design.

Table 2 shows subjects’ perception rate of the avatar’s eight facial expressions. Five of the eight emotions (anger, sadness, surprise, happiness, and neutral) received higher scores. The remaining three facial designs of the avatar’s emotions were revised according to feedback from the subjects. Interviews were conducted to gain their opinions. Seventeen subjects found it difficult to distinguish worry from sadness. In the fear countenance, 23 subjects thought the mouth opened too wide to be natural. Eight subjects thought the avatar should narrow its eyes or frown a little more when disgusted. The results were shown to our artist to improve the design of the avatar’s facial expression. Experiment 2 evaluated the perception of the avatar’s empathy where the avatar expressed empathy by her voice and upper-body performance, including facial expressions, gestures, and body movements.

| Table 2. Mean score on subjects’ perception of the virtual human’s eight expressions |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Anger                           | Sadness                         | Surprise                        | Worry                           |
| M                               | SD                              | M                               | SD                              | M                               | SD                              |
| 3.87                            | 0.84                            | 3.88                            | 0.78                            | 4.02                            | 0.94                            | 3.62                            | 0.87                            |
| Happiness                       | Disgust                         | Fear                            | Neutral                         |
| M                               | SD                              | M                               | SD                              | M                               | SD                              |
| 4.10                            | 0.77                            | 3.31                            | 1.05                            | 3.54                            | 0.94                            | 3.87                            | 0.84                            |

(1 = strongly disagree; 2 = disagree; 3 = no comment; 4 = agree; 5 = strongly agree)

Experiment 2

To investigate students’ perception of the avatar and its effect on students’ learning behavior, we exposed students to agents in controlled learning experiences to obtain students’ assessment of the avatars by questionnaires. To investigate the effects caused by the avatar, we analyzed their (1) amount of learning time, (2) willingness to continue learning, and (3) emotions in reading activities and doing exercises.

To study how different agent settings influence students’ perception and learning, we developed three settings of agents and introduced each one into a copy of the learning environment. The first setting includes the empathic avatar, which responds to participants with empathic facial expressions, voices, gestures, and body movements. The second setting is almost the same as the first setting, with only the empathic design removed from the avatar. In the third setting, the avatar performance was replaced by text. Despite the difference, the settings were identical in all other respects.

First setting: empathic avatar with empathic facial expressions, voices, gestures, and body movements.

Second setting: avatar without empathic responses and with neutral facial emotion.

Third setting: empathic responses in text without the avatar.

Participants and setting

Thirty college freshmen (24 males; 6 females) majoring in computer science participated in the evaluation. Most had been using computers for five years on average, and only a few of them had ever interacted with avatars. Students were randomly assigned to one of the three groups. Students in different groups experienced all three settings of the learning system in different order. Students in the first group used the first, second, and third setting of the system sequence. Students in the second group used the second, third, and first setting of the system sequence. Students in the third group used the third, first, and second setting of the system sequence.

The study was conducted in a meeting room. The program was installed on laptops, each with 80GB hard disc, a two-button mouse, and a 14-inch color display that ran Microsoft Windows XP operating system. The screen resolution was 1024 pixels x 768 pixels.
Materials and procedure

Each student was placed in a learning system setting with a laptop for 30 minutes on average. During this time, students read, performed exercises, interacted with an avatar, and completed a questionnaire. The avatar introduced students to the experiment and the learning system. Then students spent 5 to 15 minutes reading the e-textbook and simultaneously interacting with the avatar. Students could stop reading at any time and the amount of reading time was recorded, shown in Table 4. After reading, students entered the exercise phase. The exercise contained 40 yes/no questions. When students wanted to quit the exercise activities, the avatar persuaded them to complete the remaining questions. The number of completed questions was recorded to show differences between the three settings. After exercising, the agent encouraged students to read the e-textbook again, and their willingness in the three experimental settings was recorded for further analysis. At the end of the experiment, five-point Likert scale questionnaires and interviews were conducted.

Data analysis and results

To verify students’ perceptions of the avatar and the learning system, Table 3 lists the questionnaire results. The top lines of Table 3 present the means and standard deviations for the three settings on each of these questions. Results in question 1 showed that the empathic avatar and the empathic text on average obtained empathy transmission from the subjects, and that the former group gained higher scores than the latter one. The examination shows that students in the empathic avatar setting rate significantly higher than students in the avatar without empathy group, based on a two-tailed t test, \( t(28) = 3.08, p = 0.005 \). Students in the empathic text setting also rated significantly higher than students in the avatar without empathy group, based on a two-tailed t test, \( t(28) = 2.63, p = 0.014 \). But students in the empathic avatar setting did not rate significantly higher than students in the text-agent group, based on a two-tailed t test. Compared to the non-empathic avatar, results in question 2 showed that the empathy expressed by the avatar or in the form of text increased users’ interests. Subjective feelings also compared to their usage of the system. In question 2, the empathic avatar setting obtained the same scores as the empathic text setting, with similar results to the experiment results of the study (Moreno et al., 2001), saying that a pedagogical agent’s image in a multimedia lesson does not hurt, nor does it provide any cognitive or motivational advantage for students’ learning. Similarly, compared to the other two control groups, results in question 3 showed that the system with the empathic avatar merely had higher potential for students to use again; two-tailed t tests in question 3 showed that students in the different settings did not have significant rates.

<table>
<thead>
<tr>
<th>First setting:</th>
<th>Second setting:</th>
<th>Third setting:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empathic avatar</td>
<td>Avatar without empathy</td>
<td>Empathic text</td>
</tr>
<tr>
<td>M, SD</td>
<td>M, SD</td>
<td>M, SD</td>
</tr>
<tr>
<td>Q1: Did you sense empathy from the agent?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M = 3.73, SD = 0.80 )</td>
<td>( M = 2.86, SD = 0.74 )</td>
<td>( M = 3.53, SD = 0.64 )</td>
</tr>
<tr>
<td>Q2: Do you like the learning activities more when learning interactively with the agent?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M = 3.60, SD = 0.83 )</td>
<td>( M = 3.27, SD = 0.80 )</td>
<td>( M = 3.60, SD = 0.83 )</td>
</tr>
<tr>
<td>Q3: Are you willing to learn with the agent again?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M = 3.80, SD = 0.86 )</td>
<td>( M = 3.4, SD = 0.83 )</td>
<td>( M = 3.47, SD = 0.74 )</td>
</tr>
</tbody>
</table>

(1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree)

Besides subjective questionnaires, this research considered quantitative measures to examine the effect of encouragement and persuasion, (1) time spent on e-textbook reading and willingness to reread; (2) the completion rate of exercises and willingness to do more exercises. This study did not count time spent on interaction between the avatar and students as reading time. The e-learning system consisted of 40 exercises, and Table 4 shows the results. The empathic avatar can impel subjects to express their willingness to read more, but this does not affect their reading behavior actually. Subjects in the third setting (empathic text) spent 747.2 seconds on average in reading activity. However, in the first setting and second setting, the average number of seconds subjects spent in reading was 689.4 and 611.93, respectively. Subjects in the empathic text setting spent significantly more time reading than did students in the non-empathic avatar group, based on a two-tailed t test, \( t(28) = 3.201, p=0.003 \). However, subjects in the empathic avatar setting did not spend significantly more time reading than did subjects in the non-empathic setting. This study expected that subjects would spend much time reading, following the empathic avatar’s
persuasion. Subjects showed highest agreement to read again in the first (empathic avatar) setting. However, subjects in this setting did not spend as much time as those in the third (empathic text) setting. Subjects might have only given positive opinions in the questionnaire to satisfy the avatar’s or system developer’s expectations. The longest reading time occurring in the empathic text setting seemingly resulted from giving subjects a quiet learning environment, which helped them concentrate on what they really wanted to read. The results are similar to many studies; virtual humans might distract students and disturb their learning. The empathic avatar seemed to affect student-learning behaviors by increasing the number of exercises they did. The students in the empathic setting did not differ significantly in the number of completed exercises from students in the non-empathic setting. The results are similar to Okonkwo & Vassileva’s (2001) experimental results that students felt a need to do well to avoid disappointing or hurting the social agent’s feelings. In this study, many of them attempted to satisfy the agent’s expectations. The results support the observation that ordinary computer-literate individuals can be induced to use social rules toward computer agents and behave as if computers were human (Nass, Steuer, & Tauber, 1994).

<table>
<thead>
<tr>
<th>Activity 1: Reading e-textbook</th>
<th>Empathic avatar</th>
<th>Avatar without empathy</th>
<th>Empathic text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of the reading time (seconds)</td>
<td>$M$</td>
<td>689.40</td>
<td>611.93</td>
</tr>
<tr>
<td></td>
<td>$SD$</td>
<td>110.13</td>
<td>116.57</td>
</tr>
<tr>
<td>Willingness to read again (Percentage of participants)</td>
<td>53.33%</td>
<td>33.33%</td>
<td>40%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity 2: Doing exercises</th>
<th>Empathic avatar</th>
<th>Avatar without empathy</th>
<th>Empathic text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of answered questions</td>
<td>$M$</td>
<td>32.20</td>
<td>26.53</td>
</tr>
<tr>
<td></td>
<td>$SD$</td>
<td>8.10</td>
<td>8.25</td>
</tr>
<tr>
<td>Willing to do more exercises (Percentage of participants)</td>
<td>66.67%</td>
<td>33.33%</td>
<td>60.00%</td>
</tr>
</tbody>
</table>

In addition to the questionnaire and behavioral results, this experiment classified subjects’ emotions while the students read the e-textbook and did exercises in the three different settings. Figure 4 shows their emotions in different learning activities. The upper part of Figure 4 shows that most subjects felt interested and happy while reading. Subjects working with the empathic avatar obviously expressed their worries. To examine whether subjects in the three settings experienced different learning emotions while reading, a chi-square test was used, and results indicated that subjects’ emotions in the three settings were significantly different ($\chi^2(1) = 22.536, p = 0.032$). Subjects in the three settings were all told that their reading performance involved participating in the reading activity and doing exercises. Pressure causes worry, but only subjects in the first setting (empathic avatar) felt worry while reading. The avatar’s emotional responses and persuasion seemingly raised the participants’ feeling of worry.

![Figure 4. Subjects’ emotions while reading and doing exercises](image)
The lower part of Figure 4 shows students’ emotions while doing exercises. The students mainly felt frustrated and angry while doing exercises (answering yes/no questions) compared to when they were reading. While doing the exercises, six subjects expressed boredom in the non-empathic setting, while no one expressed boredom in the empathic setting. Compared to students in the non-empathic setting, students in the empathic-avatar setting felt worried, sad and frustrated. Students in neutral setting experienced more anger than did the students in the other two groups. To examine whether subjects in the three settings had different learning emotions while exercising, a chi-square test was used, but there were no significant differences. The result that differs from previous one might be caused by an exercise that greatly affected students’ emotions. Some participants expressed that the time constraint for answering each question caused their anger, and their ignorance during the previous reading activity caused their frustration. Far more research is needed on how students’ emotions change while learning and interacting with an avatar. Moreover, a chi-square test was used to examine whether participants have different learning emotions in reading and exercising activities. Results were statistically significant ($x^2_{(i)} = 111.656, p < 0.001$).

**Discussion and design implications**

This study designed avatars as companions that empathetically responded to users for the purposes of encouragement and persuasion. The experiment results show that an avatar who delivers empathy might affects users’ perceptions, emotions, and behavior. Compared to the non-empathic one, the empathic avatar attracts users and increases their interest, which might cause students to work harder (Harp & Mayer, 1998). Results in question 3 of Experiment 2 show that a system with an empathic avatar has higher potential for future use by students. Nevertheless, avatars might also distract students and the effect revealed in this study shows that students in the empathic-avatar group did not spend more time on reading than students in the empathic-text group. Therefore, a compromised design might be that the avatar will appear only when students need it.

The experimental results show that users felt a need to do well to satisfy the agent’s expectation in the subjective questionnaire and in actual learning behavior. Compared to the two learning activities, rereading is unspecific and intangible compared to doing more exercises, causing inconvenience for students. Persuasion seemed effective in tangible learning activities. After persuasion, subjects tended to do five to six exercises on average, compared to subjects in the non-empathic group. The willingness percentage to do more exercises in the questionnaire increased from 33.33 percent to 66.67 percent. Applying the avatar to encourage or persuade students might be more useful with concrete instructions on tangible activities. Participants felt much more frustration and anger during the second activity, when they worked more exercises according to the avatar’s persuasion.

This experiment is limited by the nature of the learning materials and few participants. Different subjects with various learning materials might determine students’ perception of the system and agent, and their learning behavior. Having participants not divided by their gender, experiences, and proficiency might affect perceptions and behavior. Exercises with different difficulty might also provoke different emotional responses. This is worth considering while designing this type of empathetic avatar in learning system.

**Conclusions**

This study proposes a methodology to create an avatar that empathically encourages and persuades students while they use e-learning systems. Students can read, do exercises, and empathically interact with an avatar at the same time. An animator created upper-body performance of the avatar to react with learners in empathic ways, trying to encourage and persuade them into persistent learning.

Results show that participators in general recognize the avatar’s emotions through facial expressions and sense empathic reactions from the avatar. Animators creating coherent voices and upper-body behavior, including facial expressions, gestures, and head and body movements of the avatar, can promote students to satisfy the avatar’s expectation, and persuade them to put forth more effort on quantitative and specific learning activities. The experiment shows that well-designed and interesting avatars can positively change learning attitudes and behaviors. The avatar’s emotional responses and persuasion might foster user’s feelings of worry, which could be utilized in an e-learning system design to enhance student’s learning performance. Moreover, different learning activities, like reading or exercising, might affect students’ emotions differently. As a result, learning activities should be
considered on a case-by-case basis while providing empathic avatars as their learning assistants. These experiments and findings could have many important implications for educational software design.

Future research could develop sophisticated avatars that clearly engage with students’ cognition and emotion. Automatic detection of emotions and cognition could also enhance functionality. The proposed methodology needs long-term examination and a large number of subjects to examine the design.

Acknowledgments

We would like to thank the researchers, teachers and students who participate in the system design, implementation and experiment. We are also grateful for the support of the National Science Council, Taiwan, under NSC 99-2631-S-008-004 and 99-2631-S-008-006-CC3. We would also like to thank the support from Research Center for Science and Technology for Learning, National Central University, Taiwan.

References


Evaluating Knowledge Structure-based Adaptive Testing Algorithms and System Development

Huey-Min Wu, Bor-Chen Kuo and Jinn-Min Yang

Research Center for Testing and Assessment, National Academy for Educational Research, New Taipei City, Taiwan // Graduate Institute of Educational Measurement and Statistic, National Taichung University of Education, Taichung, Taiwan // Department of Mathematics Education, National Taichung University of Education, Taichung, Taiwan // lhswu@seed.net.tw // kbc@mail.ntcu.edu.tw // ygm@ms3.ntcu.edu.tw

(Submitted September 11, 2009; Revised January 26, 2011; Accepted March 31, 2011)

ABSTRACT

In recent years, many computerized test systems have been developed for diagnosing students’ learning profiles. Nevertheless, it remains a challenging issue to find an adaptive testing algorithm to both shorten testing time and precisely diagnose the knowledge status of students. In order to find a suitable algorithm, four adaptive testing algorithms based on ordering theory, item relational structure theory, Diagnosys, and domain experts, were evaluated based on the training sample size, prediction accuracy, and the use of test items by the simulation study with paper-based test data. Based on the results of simulation study, ordering theory has the best performance. An ordering-theory-based knowledge-structure-adaptive testing system was developed and evaluated. The results of this system showed that the two different interfaces, paper-based and computer-based, did not affect the examinees’ performance. In addition, the effect of correct guessing was discussed, and two methods with adaptive testing algorithms were proposed to mitigate this effect. The experimental results showed that the proposed methods improve the effect of correct guessing.

Keywords
Adaptive test algorithm, Computerized adaptive test, Diagnostic test, Knowledge structure, Ordering theory

Introduction

During the last two decades, from the functional aspect, many computerized test systems have been developed for estimating abilities of examinees (Chang, Lin, & Lin, 2007; Guzman & Conejo, 2005; Lewis & Sheehan, 1990; Sands, Water, & McBride, 1997; Sheehan & Lewis, 1992; Wainer, 2000; van der Linden, 2000; Tao, Wu, & Chang, 2008; Yen, Ho, Chen, Chou, or Chen, 2010) or diagnosing students’ learning profiles (Appleby, Samuels, & Treasure-Jones, 1997; Chang, Liu, & Chen, 1998; Hwang, Hsiao, & Tseng, 2003; Liu, 2005; Tsai & Chou, 2002; Tselios, Stoica, Maragoudakis, Avouris, & Komis, 2006; Vomlel, 2004; Yu & Yu, 2006). From the theoretical aspect, some of them are based on item-response theory (IRT) (Chang et al., 2007; Guzman & Conejo, 2005; Lewis & Sheehan, 1990; Sands et al., 1997; Sheehan & Lewis, 1992; Wainer, 2000; van der Linden, 2000; Yen et al., 2010), some of them are based on artificial intelligence techniques such as Bayesian networks (Liu, 2005; Tselios et al., 2006; Vomlel, 2004), and others are based on knowledge structures. From the operational aspect, some of the computerized tests are adaptive and others are non-adaptive. The focus of this study is to construct computerized adaptive tests based on knowledge structures for diagnosing students’ learning profiles.

The computerized adaptive test (CAT) can not only offer examinees customized items in accordance with their aptitudes or cognitive status, but can also shorten the test. The CAT based on IRT models can obtain efficient estimates of subjects’ abilities, but it cannot provide the capability to diagnose subjects’ cognitive concepts at a detailed level (Tatsuoka, Corter, & Tatsuoka, 2004; Yan, Almond, & Mislevy, 2004). Instead, knowledge structure-or artificial-intelligence-based adaptive tests can provide information about how well subjects performed on specific concepts, so they can achieve the diagnostic function (Appleby et al., 1997; Tatsuoka et al., 2004; Vomlel, 2004).

Diagnosys, developed by Appleby et al. (1997), is a knowledge-based-computer diagnostic test of basic mathematical concepts. In Diagnosys, a method was proposed to estimate the knowledge structure of examinees and then apply this structure to build the adaptive testing process. Chang et al. (1998) have proposed adaptive test algorithms to construct a computerized adaptive diagnostic test based on knowledge structures constructed by the domain experts. The results of these two papers exhibit that the proposed algorithms have the capability of decreasing the use of test items and are able to precisely diagnose the cognitive status of examinees. However, the impact of correct guessing on the diagnoses of concepts is not considered in these studies. Correct guessing means...
that an item is answered correctly by guessing in multiple-choice tests. In knowledge-based adaptive tests, if an item is answered correctly by guessing, then all prerequisite items of it are assumed to have been answered correctly. But, in actual fact, these prerequisite items may not have been answered correctly. In that situation, the precision of diagnosing results would be decreased. Moreover, the impact of correct guessing in adaptive testing would be greater than that in non-adaptive testing such as the traditional paper-and-pencil test.

Tselios et al. (2006) used the Bayesian network to diagnose students’ problem-solving strategies with two distinct problems. The results show that the Bayesian network can estimate students’ problem-solving strategies very well, but it is not an adaptive test. Vomlel (2004) and Liu (2005) have proposed adaptive testing algorithms based on the Bayesian network. In their simulation study, the numbers of test items were 10 and 21, respectively. The experimental results show that the Bayesian network is a powerful tool to diagnose students’ learning status; however, it is difficult and time consuming to find the optimal adaptive testing strategy when the test is long.

Ordering theory (OT; Airasian & Bart, 1973; Bart & Krus, 1973) and item relational structure theory (IRS; Takeya, 1991) were proposed for displaying the students’ item structures. In previous studies (Bart & Krus, 1973; Takeya, 1991), OT and IRS were used for developing instruction sequences or learning progress indices. In this paper, OT and IRS are used to estimate knowledge structures of examinees and apply them to new adaptive test algorithms. One of the currently existing problems is that there are many knowledge-structure-based adaptive testing (KSAT) algorithms but no study to evaluate their performance. The performance of the adaptive testing algorithm, Diagnosys, domain experts, OT, and IRS is evaluated by using the simulation study; moreover, the effect of correct guessing in the multiple-choice tests are also explored in this study. In comparison to the algorithm proposed by Appleby et al. (1997) and the domain experts, our algorithms significantly reduce the length of time to take tests, and the algorithm with the best performance is selected to construct a computerized adaptive diagnostic test to be used in an actual Grade five diagnostic mathematics test. The experimental results show that the computerized adaptive diagnostic test has performed as expected.

Adaptive test algorithms based on knowledge structures

A hierarchy concept network, knowledge structure, introduced by Gagne (1977) as a way of defining prerequisite association of concepts, is the combination of named individual concepts, a specified level for each concept, and specified directed links between concepts that joins them together into a hierarchy. As shown in Figure 1, concept D is linked forwardly to concept C, which means that concept D must be mastered before concept C can be attempted; that is, concept D is a prerequisite for concept C.

By using this concept network, Appleby et al. (1997) proposed an inference mechanism (adaptive testing algorithm) that allowed the system to reduce the number of items that are administered in computerized adaptive diagnostic test. As shown in Figure 1, if the student gets concept D correct then it is inferred that he or she also knows its prerequisites (concepts F, G, H, and I). This algorithm in computerized adaptive diagnostic test can predict students’ learning profiles by using fewer items than original paper-based tests.

The number of links has an impact on the use of test items. As the number of links increases, the use of test items decreases. In this paper, we propose adaptive testing algorithms with OT and IRS respectively.

![Figure 1. The knowledge structure](image-url)
The domain experts’ knowledge structure

Once a knowledge structure is constructed by practising teachers and domain experts, it is named as the domain experts’ knowledge structure. The procedures for constructing a domain experts’ knowledge structure are as follows. First, the domain experts define the important concepts of each unit by analyzing teaching materials and objectives. Second, after much discussion, the domain experts decide the sequence of the concepts development and relationships among these concepts to depict in a tree diagram the experts’ knowledge structure for each unit. Figure 2 is an example of part of the domain experts’ knowledge structure for a triangle unit of mathematics used in elementary schools of Taiwan. In the domain experts’ knowledge structure, the upper-level concepts such as “find the isosceles triangle” are advanced concepts, while low-level concepts such as “find the right angle” are basic-level concepts. Generally, an item is developed to assess knowledge on a single concept. Diagnostic tests are developed by the concepts defined in the domain experts’ knowledge structure.

[Diagram of the domain experts’ knowledge structure for the triangle]

Figure 2. Part of the domain experts’ knowledge structure for the triangle

Knowledge structure from Diagnosys

With Diagnosys, a paper-based pre-test is developed based on domain experts’ structures and is then administered to collect responses from students. This data was applied to develop the inference mechanism as follows.

The relative frequencies of two concepts, $A$ and $B$, are defined in Table 1. As shown in Table 1, $f_{AB}$ represents the number of students with correct answers for both concept $A$ and concept $B$. If $f_{AB} + f_{AB} > f_{AB} + f_{AB}$, then concepts $A$ and $B$ are equivalent and the relation is denoted as $A \leftrightarrow B$. Therefore, if students understand concept $A$, they will understand concept $B$ as well, and vice versa. Moreover, if $f_{AB} > f_{AB}$, then concept $A$ could be linked forwards to concept $B$. The relation denoted as $A \rightarrow B$ means that $A$ is a prerequisite to $B$. The important characteristic of the link $A \rightarrow B$ is twofold:

1. If the student gets an item on concept $B$ correct, we can infer that she or he also understands concept $A$.
2. If the student gets an item on concept $A$ incorrect, we can infer that she or he also does not understand concept $B$.

These two rules apply transitively across the structure according to the partial ordering given by the links. For example, for the network, $A \rightarrow B \rightarrow C$, if a student gets an item on concept $C$ correct then we can infer that the student understands concept $B$ due to direct inferences, but also $A$ due to indirect transitive inferences. This algorithm allows the system to significantly reduce the number of items administered compared with a conventional test.
Some inefficient problems are posed such as the definitions of ordering relation, equivalence and transition among concepts lack clarity while being inoperable. To improve these limitations, a threshold model is defined in this paper: If \( \eta_{AB}^* = \frac{f_{\overline{A}B} + f_{AB}}{f_{\overline{A}B} + f_{AB}} < \epsilon_{dia} \) then \( A \leftrightarrow B \).
If \( \lambda_{AB}^* = f_{AB}/f_{\overline{A}B} < \epsilon_{dia} \), then \( A \rightarrow B \).

### Ordering theory and item relational structure theory

In this paper, two other item ordering theories, OT and IRS are used for estimating knowledge structures of examinees and to develop new adaptive test algorithms. They are described briefly below:

Let \( X = (X_1, X_2, \ldots, X_n) \) denote a vector containing \( n \) binary item score variables. Each student taking an \( n \)-item test produces a vector \( X = (x_1, x_2, \ldots, x_n) \) containing 1 (correct) and 0 (incorrect). Then the joint and marginal probabilities of items on concepts A and are represented in Table 2.

<table>
<thead>
<tr>
<th>concept B</th>
<th>( X_B = 1 )</th>
<th>( X_B = 0 )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_A = 1 )</td>
<td>( P(X_A = 1, X_B = 1) )</td>
<td>( P(X_A = 1, X_B = 0) )</td>
<td>( P(X_A = 1) )</td>
</tr>
<tr>
<td>( X_A = 0 )</td>
<td>( P(X_A = 0, X_B = 1) )</td>
<td>( P(X_A = 0, X_B = 0) )</td>
<td>( P(X_A = 0) )</td>
</tr>
<tr>
<td>Total</td>
<td>( P(X_B = 1) )</td>
<td>( P(X_B = 0) )</td>
<td>1</td>
</tr>
</tbody>
</table>

For OT, let \( \epsilon_{AB}^* = P(X_A = 0, X_B = 1) < \epsilon_{OT} \), usually \( 0.02 < \epsilon_{OT} < 0.04 \) (Airasian & Bart, 1973; Bart & Krus, 1973), concept A can be linked forward to concept B. The relation is denoted as \( A \rightarrow B \) this means that A is a prerequisite to B. If \( A \rightarrow B \) and \( B \rightarrow A \), then the relation is denoted as \( A \leftrightarrow B \) and it means concepts A and B are equivalent.

For IRS, Takeya (1991) proposed another index, \( r_{AB}^* \), which is used to define the ordering relation from concept A to concept B. The definition of \( r_{AB}^* \) is
\[
r_{AB}^* = 1 - \frac{P(X_A = 0, X_B = 1)}{P(X_A = 0)P(X_B = 1)} \geq \epsilon_{IRS}
\]
If \( r_{AB}^* \geq \epsilon_{IRS} \), then concept A can be linked forward to concept B. Usually the rule of thumb is to set \( \epsilon_{IRS} = 0.5 \).

### The performances of knowledge-structure-based adaptive testing (KSAT) algorithms

As mentioned above, four methods, Diagnosys, OT, IRS, and the domain experts, can be used to define knowledge structures. By applying these knowledge structures, the corresponding inference mechanisms (adaptive testing algorithm) are established. In this paper, we refer to them as knowledge-structure-based adaptive testing (KSAT). In this section, the performances of adaptive testing algorithms based on the four knowledge structures with different thresholds are compared and evaluated by using adaptive test simulation processes with a paper-based test dataset to determine the best algorithm. In these simulation processes, a paper-based test is taken owing to a limitation of computer equipment. The reason for using simulation is that there are hundreds of combinations of knowledge structure-based adaptive testing (KSAT) algorithms and thresholds. Finding a real computerized dataset for each combination is not feasible.
The use of test items and prediction accuracy of each combination are considered its performance. Mathematics definitions are defined in Table 3. As shown in Table 3, $f_{ij}^{11}$ represents the frequency with which student $j$ answered item $i$ correctly, both in the simulated computerized adaptive diagnostic test and in the paper-based test; $f_{ij}^{00}$ represents the frequency with which student $j$ answered item $i$ incorrectly, both in the simulated computerized adaptive diagnostic test and in the paper-based test. The prediction accuracy reflects a degree of similarity in the examinee’s responses to the simulated computerized adaptive diagnostic test and the paper-based test. The use of test items is the average items administered to the examinees in the computerized adaptive diagnostic test. One of the goals of this paper is to find the best algorithm, which is able to achieve better prediction accuracy with fewer averaged use of test items. Once the best algorithm is determined by the training data, it is used in the actual computerized adaptive test.

<table>
<thead>
<tr>
<th>Table 3. Definition of prediction accuracy and utilization of test items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated computerized adaptive diagnostic test</td>
</tr>
<tr>
<td>Correct (1)</td>
</tr>
<tr>
<td>Incorrect (0)</td>
</tr>
<tr>
<td>Paper-based test</td>
</tr>
<tr>
<td>Correct (1)</td>
</tr>
<tr>
<td>Incorrect (0)</td>
</tr>
<tr>
<td>Prediction accuracy: $PA_e = \frac{1}{N \cdot n} \sum_{j=1}^{N} \sum_{i=1}^{n} (f_{ij}^{11} + f_{ij}^{00})$ where</td>
</tr>
<tr>
<td>$\varepsilon$ : threshold, $\varepsilon = 0.01, \ldots, 0.5$ for Diagnosys and OT; $\varepsilon = 0.02, \ldots, 1$ for IRS</td>
</tr>
<tr>
<td>$j$ : the examinee from test samples $j = 1, 2, \ldots, N$</td>
</tr>
<tr>
<td>$i$ : the item $i = 1, 2, \ldots, n$</td>
</tr>
<tr>
<td>Use of test items: $UI_e = \frac{1}{N} \sum_{j=1}^{N} n_j$ where</td>
</tr>
<tr>
<td>$\varepsilon$ : threshold, $\varepsilon = 0.01, \ldots, 0.5$ for Diagnosys and OT; $\varepsilon = 0.02, \ldots, 1$ for IRS</td>
</tr>
<tr>
<td>$n_j$ : the number of items that are administered to the examinee $j$ in the computerized adaptive diagnostic test</td>
</tr>
<tr>
<td>$j$ : the examinee from test samples $j = 1, 2, \ldots, N$</td>
</tr>
</tbody>
</table>

To take an example from Figure 1, if a student has completed a paper-based test consisting of nine items, the response patterns are shown in Table 4. In the simulation KSAT process, if the student gets concept D correct then we can infer that the student also understood concepts F, G, H, and I, although they were not administered. Compared to the responses of the paper-based test, prediction accuracy and utilization of test items are calculated by the above-mentioned formula, $PA_e = \frac{8}{9} = 0.89$, $UI_e = 5$.

<table>
<thead>
<tr>
<th>Table 4. Responses for a paper-based test and a simulated computerized adaptive diagnostic test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student responses</td>
</tr>
<tr>
<td>Paper-based test</td>
</tr>
<tr>
<td>Simulated computerized adaptive diagnostic test</td>
</tr>
</tbody>
</table>

0: incorrect 1: correct 1: inferred correct

**Implementation of knowledge structure-based adaptive test (KSAT) system**

The knowledge structure-based adaptive testing (KSAT) system has been implemented with PHP and MySQL on APACHE web servers. Figure 3 shows the architecture of the knowledge structure-based adaptive testing (KSAT), which consists of 10 modules: Account Management Module, Item Bank Management Module, Test Management Module, Competency Module, Diagnosis Module, Adaptive Item Selection Module, User-profile Database, Item Bank Database, Knowledge Structure Database, and Test Result Database.
The Account Management Module provides creation and management of user accounts. The functions of Item Bank Management Module include items or the knowledge structure of specific unit updates, modification, and management. The Test Management Module can set the approach of test administration. The Competency Module estimates the competency of individual students or groups. The Diagnosis Module diagnoses the knowledge states of the student by using the response pattern of the student. The Adaptive Item Selection Module can administrate tests according to different adaptive test algorithms. According to the experiment results, the knowledge structure estimated by the ordering theory has been used to construct the adaptive test algorithm that was placed in this module.

The following are several major interfaces of system.

The user management interface in Figure 4 is multi-functional. It allows new users to have access to creating new user accounts, creating multiple new user accounts, importing accounts from other sources such as Excel, and giving access to the database.

The test administration interface in Figure 5 displays the items and allows the examinees to answer the items presented. Since the KSAT system is an adaptive test, only one item per screen is presented.

The group profile interface in Figure 6 displays the group result of the exam. For example, in concept 5 of the interface, 13 students passed and 19 students failed test 1. Instructors can then take this information and understand the distribution of students’ knowledge states and identify the strengths and weaknesses within a group. This information can be utilized for remedial instruction.

The individual profile interface is shown in Figure 7 and 8. Upon completion of the test, the student receives a personalized profile including name, scores, percentile, utilization of test items, date taken, and so forth. In Figure 14, the competency of the student for each concept in forms 1 to 3 is displayed.

Figure 3. Architecture of knowledge structure-based adaptive testing (KSAT) system
User Management

Create a new user account

* ID:

* Password:

Re-type Password:

* Name:

* Gender:
  - male
  - female

* Student:

* Birthday:

An ID number:

Telephone:

Cell phon:

Address:

Email:

Create

* required item

Figure 4. The user management interface

How many obtuse triangles are shown in the below figure?

○ 4
○ 5
○ 6
○ 7

Next item

Figure 5. The test administration interface

The Group Profile

Select class and test form:

Taiwanese test

* Option 1: Show the statistics for the class only
* Option 2: Download .csv file

Taiwanda High School, City 4 Class 5 Diagnostic Profile

<table>
<thead>
<tr>
<th>Concepts list</th>
<th>Diagnostic profile</th>
<th>Form 1 (pass/fail)</th>
<th>Form 2 (pass/fail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept 1: Recognize the right angle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 2: Recognize the isosceles triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 3: Recognize the acute angle</td>
<td>25-03</td>
<td>25-03</td>
<td></td>
</tr>
<tr>
<td>Concept 4: Recognize the obtuse angle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 5: Recognize the isosceles triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 6: The interior angles of a triangle add up to 180</td>
<td>30-00</td>
<td>30-00</td>
<td></td>
</tr>
<tr>
<td>Concept 7: Recognize the acute angle</td>
<td>20-03</td>
<td>20-03</td>
<td></td>
</tr>
<tr>
<td>Concept 8: Recognize the obtuse angle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 9: Recognize the isosceles right triangle</td>
<td>15-15</td>
<td>15-15</td>
<td></td>
</tr>
<tr>
<td>Concept 10: Test the isosceles triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 11: Calculate the unknown angle with the properties of isosceles triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 12: Recognize the scalene triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 13: Recognize the obtuse triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
<tr>
<td>Concept 14: Test the scalene triangle</td>
<td>32-00</td>
<td>32-00</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. The interface for groups profiles
Experiment 1 and results

The triangle unit of mathematics used in elementary schools of Taiwan was adopted to develop a paper-based test consisting of 35 items. The triangle mathematic test was administered to 660 selected fifth-grade students. As noted previously, four methods to define knowledge structures are mentioned. Three of the four methods require thresholds, ε, whereas the domain expert’s structure does not require a threshold. The threshold effects of three algorithms (Diagnosys, OT, and IRS) on the prediction accuracy and use of test items were explored in this experiment. The responses of selected students were randomly divided into two parts, training samples and test samples. The training samples were applied to estimate the knowledge structures, and the test samples were used to estimate the prediction accuracy and use of test items. This process was repeated 50 times to obtain 50 sets of prediction accuracy and the use of test items. The averages of prediction accuracy and use of test items were used to represent the algorithm performance. The standard deviations of prediction accuracy and use of test items were used to evaluate the stability of the four algorithms. Training samples (TS) 10, 50, 100, and 200 were used to investigate the impact of the sample size on the prediction accuracy and on the use of test items of different algorithms.
Figures 9 to 18 present the prediction accuracy and the use of test items of different adaptive testing algorithms with different training sample (10, 50, 100, and 200). The scale of the horizontal axis of IRS (thresholds 0.02, 0.04, . . . 0.98) is different from those of Diagnosys and OT (thresholds 0.01, 0.02, . . . 0.50), so it is not displayed in the same graph. The horizontal axis represents the threshold, $\varepsilon$, and the vertical axis represents the prediction accuracy, $PA_\varepsilon$ (Figures 9, 11, 13, 15, and 17) and the use of test items, $UI_\varepsilon$ (Figures 10, 12, 14, 16 and 18). For example, in Figures 9 and 10, if the knowledge structure of Diagnosys with $\varepsilon = 0.08$ was applied, then $(PA_{0.08}, UI_{0.08}) = (0.821,1.20)$, under the training samples, $(TS) = 10$. In Figures 17 and 18, if the knowledge structure of IRS with $\varepsilon = 0.58$ was applied, then $(PA_{0.58}, UI_{0.58}) = (0.896,6.35),(0.952,14.33),(0.970,19.12),(0.984,23.77)$ under the training samples (TS) = 10, 50, 100, and 200, respectively. The prediction accuracy and the use of test items of the algorithm based on domain experts’ structure are 0.917 and 18, respectively. Since constructing the domain experts’ structure does not need thresholds, it does not vary by thresholds.

Those figures show that:

1. Overall, the prediction accuracy and use of test items of Diagnosys and OT increase as the threshold decreases. The prediction accuracy and use of test items of IRS increase as the threshold increases.
2. Compared with the results from domain expert’s structure ( $(PA, UI) = (0.917,18)$ ), IRS $(PA_{0.32}, UI_{0.32}) = (0.923,8.58)$, and OT $(PA_{0.08}, UI_{0.08}) = (0.922,8.37)$) are able to achieve higher prediction accuracies with less use of test items.
3. For three test adaptive algorithms, Diagnosys, OT and IRS, the Diagnosys requires more training samples and higher use of test items to achieve the same or almost the same prediction accuracy in comparison to OT and IRS.
4. The performance of OT is less sensitive to the training sample size than that of IRS and Diagnosys.

For reducing the paper length without loss the generality, only three cases (case 1, case 2, and case 3) of means and standard deviations of prediction accuracies and their corresponding use of test items under the training sample size 200 are displayed in Table 5. Case 1, case 2, and case 3 mean the prediction accuracy, 0.90, 0.92, and 0.94, respectively. The reason for choosing these cases in range of 0.90 to 0.94 is that this range is around the prediction accuracy of domain experts’ structure and 0.94 is the maximum prediction accuracy that Diagnosys can achieve.

For example, in Diagnosys, when the average of prediction accuracy and its corresponding use of test items are 0.90 and 25.68, respectively, the standard deviations are 0.023 and 5.67, respectively. The range of standard deviations for prediction accuracy is 0.004 to 0.023, indicating that this simulation model is reliable. The lowest standard deviations of the prediction accuracy and the use of test items are all for the OT, so the OT has better performance on stability.

According to the results of the experiment, Diagnosys requires a large sample size and a larger use of test items to obtain better prediction accuracy, so it is not suggested for use. OT can obtain better prediction accuracy with less use of test items and training samples; hence OT is implemented into the KSAT system.

![Figure 9. The prediction accuracy of Diagnosys, OT, and expert for training samples 10](image)
Figure 10. The use of test items of Diagnosys, OT, and expert for training samples 10

Figure 11. The prediction accuracy of Diagnosys, OT, and expert for training samples 50

Figure 12. The use of test items of Diagnosys, OT, and expert for training samples 50

Figure 13. The prediction accuracy of Diagnosys, OT, and expert for training samples 100
Figure 14. The use of test items of Diagnosys, OT, and expert for training samples 100

Figure 15. The prediction accuracy of Diagnosys, OT, and expert for training samples 200

Figure 16. The use of test items of Diagnosys, OT, and expert for training samples 200

Figure 17. The prediction accuracy of the IRS for training samples 10, 50, 100, and 200
Table 5. The means and standard deviations (in brackets) of the prediction accuracy and use of test items

<table>
<thead>
<tr>
<th>Case</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prediction accuracy</td>
<td>0.90 (0.023)</td>
<td>0.92 (0.015)</td>
</tr>
<tr>
<td>use of test items</td>
<td>25.68 (5.67)</td>
<td>28.89 (2.43)</td>
</tr>
<tr>
<td>IRS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prediction accuracy</td>
<td>0.90 (0.010)</td>
<td>0.92 (0.007)</td>
</tr>
<tr>
<td>use of test items</td>
<td>5.94 (1.003)</td>
<td>8.57 (1.163)</td>
</tr>
<tr>
<td>OT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prediction accuracy</td>
<td>0.90 (0.004)</td>
<td>0.92 (0.004)</td>
</tr>
<tr>
<td>use of test items</td>
<td>5.60 (0.400)</td>
<td>8.37 (0.580)</td>
</tr>
</tbody>
</table>

Improvement of the correct guessing in KSAT algorithms

There are two major factors that affect the performances of adaptive testing algorithms. One is the theory to build knowledge structures and the other is the correct guessing of multiple-choice items. The effects of different theories are shown in the experiment 1, and we will discuss how to reduce the effect of the correct guessing in this section. In KSAT algorithms, if an item is answered correctly by guessing then all the prerequisite items of it will be assumed to be correct answers. This correct guessing would decrease the prediction accuracy of KSAT algorithms. Actually, the statistical nature of KSAT algorithms (especially OT) has the function to reduce the effect of correct guessing. Take OT as an example. OT, \( \epsilon_{AB} = P(X_A = 0, X_B = 1) < \epsilon_{OT} \), if \( P(X_A = 0, X_B = 1) \geq \epsilon_{OT} \) (i.e., concept A is not a prerequisite of concept B), then the correct guessing only affects the prediction accuracy of concept B; otherwise the prediction accuracy of concept A will be influenced by the correct guessing. If the threshold is small, then the effect of the correct guessing decreases. But the use of test items will increase.

To improve this situation, two methods are proposed with KSAT. Take Figure 1 as an example, these two methods are described in the following.

Most difficult item (MDI) method: Suppose item C is answered correctly, then the most difficult item (suppose this is item B) in its prerequisite items will be presented to the examinee. If item B is answered correctly, then item C and its prerequisite items are recorded correct; otherwise, C is recorded and other prerequisite items should be taken by the examinee.

Prerequisite Item method (PI method): If item C is answered correctly, then the item with the largest number of prerequisite items (for example, item D) in C’s prerequisite items will be presented to the examinee. If item D is answered correctly, then item C and its prerequisite items are recorded as correct; otherwise, C is recorded as incorrect and other prerequisite items should be taken by the examinee. If none of the prerequisite items of C have a prerequisite item, then a randomly selected item is applied to the examinee.
## Experiment 2 and results

In this experiment, the performance of nine adaptive testing algorithms, Diagnosys, Diagnosys+MDI, Diagnosys+PI, OT, OT+MDI, OT+PI, IRS, IRS+MDI, and IRS+PI, were evaluated by using the same data set as Experiment 1. The use of test items and prediction accuracy were obtained by five-fold cross-validation. Results were presented in Table 6. For example, in Table 6, when the threshold was 0.01, the prediction accuracy of Diagnosys, Diagnosys+MDI, and Diagnosys+PI were 0.956, 0.996, and 0.992, respectively, and their corresponding use of test items were 32.64, 34.70, and 34.53. A Wilcoxon-Signed-Ranks test was used to compare the performances among nine models (Table 7). In Table 7, “Diag+MDI to Diag” indicates that the performance between original Diagnosys and Diagnosys with MDI was compared. Due to different thresholds, the performance of Diag+MDI, OT+MDI, and IRS+MDI were not explored. The results are as follows.

1. In Table 7, the results of the Wilcoxon-signed-ranks test revealed that Diagnosys, OT, and IRS, the prediction accuracies, adaptive testing algorithms with the most-difficult-item method (MDI) and prerequisite method (PI) both perform better than the original adaptive testing algorithms ($z = -3.422 \sim -3.409, p = 0.001$). Otherwise, in Diagnosys, Diagnosys+MDI outperform Diagnosys+PI ($z = -3.415, p = 0.001$); in OT, OT+MDI outperform OT+PI ($z = -3.066, p = 0.002$); in IRS, IRS+MDI outperformed IRS+PI ($z = -3.482, p = 0.000$). Overall, the performance of adaptive testing algorithms with the most-difficult item (MDI) method was better than that of adaptive testing algorithms with the prerequisite method (PI method).

2. In Table 6, under the same (or almost the same) prediction accuracies, the use of test items in the proposed KSAT+MDI and KSAT+PI are fewer than those in the original KSAT algorithms. For example, in Table 6, when the prediction accuracies of Diagnosys, Diagnosys+MDI, and Diagnosys+PI are 0.945, 0.943, and 0.945, respectively and their corresponding use of test items are 31.85, 24.23, and 30.81. When the prediction accuracy of OT, OT+MDI, and OT+PI are 0.991, 0.991, and 0.991, respectively, their corresponding use of test items are 27.27, 26.25, and 26.18. When the prediction accuracy of IRS, IRS+MDI, and IRS+PI are 0.991, 0.991, and 0.991, respectively, their corresponding use of test items are 27.75, 26.38, and 27.47 (see grayed cells).

3. In Table 6, OT+MDI and OT+PI outperform Diagnosys+MDI, Diagnosys+PI, IRS+MDI, and IRS+PI at the same prediction accuracies. The only exception is at the prediction accuracy of 0.997, where the use of test items of OT+PI and IRS_MDI are 31.46 and 30.82, respectively (see bold cells).

### Table 6. The prediction accuracy and use of test items (in brackets) of Diagnosys, OT, and IRS with MDI or PI

<table>
<thead>
<tr>
<th>Diag threshold</th>
<th>Diag +MDI</th>
<th>Diag +PI</th>
<th>OT threshold</th>
<th>OT +MDI</th>
<th>OT +PI</th>
<th>IRS threshold</th>
<th>IRS +MDI</th>
<th>IRS +PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.956</td>
<td>0.996</td>
<td>0.992</td>
<td>0.01</td>
<td>0.995</td>
<td>0.997</td>
<td>0.58</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>(32.64)</td>
<td>(34.70)</td>
<td>(34.53)</td>
<td></td>
<td>(29.98)</td>
<td>(31.51)</td>
<td></td>
<td>(31.46)</td>
</tr>
<tr>
<td>0.015</td>
<td>0.945</td>
<td>0.993</td>
<td>0.992</td>
<td>0.015</td>
<td>0.991</td>
<td>0.996</td>
<td>0.56</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>(31.85)</td>
<td>(34.50)</td>
<td>(34.45)</td>
<td></td>
<td>(27.27)</td>
<td>(29.18)</td>
<td></td>
<td>(29.16)</td>
</tr>
<tr>
<td>0.02</td>
<td>0.935</td>
<td>0.990</td>
<td>0.985</td>
<td>0.02</td>
<td>0.985</td>
<td>0.991</td>
<td>0.54</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>(31.02)</td>
<td>(34.25)</td>
<td>(34.05)</td>
<td></td>
<td>(24.41)</td>
<td>(26.25)</td>
<td></td>
<td>(26.18)</td>
</tr>
<tr>
<td>0.025</td>
<td>0.927</td>
<td>0.986</td>
<td>0.978</td>
<td>0.025</td>
<td>0.979</td>
<td>0.987</td>
<td>0.52</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>(30.22)</td>
<td>(33.92)</td>
<td>(33.58)</td>
<td></td>
<td>(22.14)</td>
<td>(24.07)</td>
<td></td>
<td>(23.94)</td>
</tr>
<tr>
<td>0.03</td>
<td>0.918</td>
<td>0.983</td>
<td>0.969</td>
<td>0.03</td>
<td>0.972</td>
<td>0.981</td>
<td>0.5</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>(29.31)</td>
<td>(33.50)</td>
<td>(32.93)</td>
<td></td>
<td>(19.42)</td>
<td>(21.59)</td>
<td></td>
<td>(21.42)</td>
</tr>
<tr>
<td>0.035</td>
<td>0.920</td>
<td>0.982</td>
<td>0.966</td>
<td>0.035</td>
<td>0.966</td>
<td>0.976</td>
<td>0.48</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td>(29.50)</td>
<td>(33.43)</td>
<td>(32.62)</td>
<td></td>
<td>(17.06)</td>
<td>(19.13)</td>
<td></td>
<td>(19.01)</td>
</tr>
<tr>
<td>0.04</td>
<td>0.912</td>
<td>0.975</td>
<td>0.945</td>
<td>0.04</td>
<td>0.960</td>
<td>0.972</td>
<td>0.46</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>(28.14)</td>
<td>(32.43)</td>
<td>(30.81)</td>
<td></td>
<td>(15.59)</td>
<td>(17.68)</td>
<td></td>
<td>(17.43)</td>
</tr>
<tr>
<td>0.045</td>
<td>0.908</td>
<td>0.972</td>
<td>0.928</td>
<td>0.045</td>
<td>0.955</td>
<td>0.967</td>
<td>0.44</td>
<td>0.967</td>
</tr>
<tr>
<td></td>
<td>(27.20)</td>
<td>(31.83)</td>
<td>(29.31)</td>
<td></td>
<td>(14.50)</td>
<td>(16.54)</td>
<td></td>
<td>(16.32)</td>
</tr>
<tr>
<td>0.05</td>
<td>0.896</td>
<td>0.958</td>
<td>0.903</td>
<td>0.05</td>
<td>0.949</td>
<td>0.962</td>
<td>0.42</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>(23.55)</td>
<td>(28.52)</td>
<td>(24.97)</td>
<td></td>
<td>(13.30)</td>
<td>(15.43)</td>
<td></td>
<td>(15.27)</td>
</tr>
<tr>
<td>0.055</td>
<td>0.893</td>
<td>0.954</td>
<td>0.899</td>
<td>0.055</td>
<td>0.943</td>
<td>0.958</td>
<td>0.4</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>(22.40)</td>
<td>(27.36)</td>
<td>(23.76)</td>
<td></td>
<td>(11.95)</td>
<td>(14.23)</td>
<td></td>
<td>(14.06)</td>
</tr>
<tr>
<td>0.06</td>
<td>0.883</td>
<td>0.943</td>
<td>0.886</td>
<td>0.06</td>
<td>0.938</td>
<td>0.954</td>
<td>0.38</td>
<td>0.949</td>
</tr>
</tbody>
</table>
Evaluation of the KSAT system

To evaluate the performance of the KSAT system, an experiment has been conducted. This experiment aimed to evaluate the efficiencies of use of test items and prediction accuracy in administering the computerized adaptive test. One hundred and twenty-three students from fifth-grade classes of Taiwanese elementary schools participated in this experiment. The procedure was conducted as follows. First, all students received a knowledge-structure-based adaptive testing (KSAT) based on OT algorithm (threshold = 0.05). The content of the test was on the triangle unit, as mentioned above. Then, when the students completed the adaptive portion of the test, the system administered the rest of the 35 items in order to compute the prediction accuracy. Finally, the use of test items and prediction accuracy were computed.

After completion of the test, the results of the use of test items and prediction accuracy were 11.42% and 93%, respectively. The results show that the KSAT system can decrease the use of test items and are able to precisely diagnose the cognitive status of examinees.

These results are consistent with the results of OT case in the previous simulation experiment (the use of test items: 13; prediction accuracy: 95%). The two different interfaces, paper-based and computer-based, do not affect the examinees’ performance in adaptive tests.

For exploring the performances of OT+MDI and OT+PI in this system, since the responses of all 35 items were available, this data set was applied to simulate OT+MDI and OT+PI processes. This simulated result shows that the use of test items and prediction accuracy of OT+MDI and OT+PI processes were (11.3, 94%) and (11.2, 94%), respectively. This implies OT+MDI and OT+PI have better performance than original OT, which is similar to the results of experiment 2.

Discussions and conclusions

In this paper, some traditional item ordering theories (OT and IRS) that were used to develop instruction sequences or learning progress indices were applied to develop the computerized adaptive testing processes. The performances of the adaptive testing algorithms based on the item structures constructed by OT, IRS, Diagnosys, and domain experts were evaluated. Three findings were found from the experimental results. First, OT and IRS based KSAT algorithms provide better prediction accuracy with less the use of test items. Second, OT-based KSAT algorithm is less sensitive to the training sample size. Third, the estimation error of OT method is less than others and this means that the diagnostic results estimated by OT-based KSAT is more stable. From the theoretic view of OT,
\( \varepsilon_{AB}^* = P(X_A = 0, X_B = 1) \) is the probability of violating the ordering relationship of \( A \rightarrow B \); that is directly related to the prediction error. From the definition and explanation of IRS in Takeya (1991), \( r_{AB}^* \) is designed to be a coefficient that has the benefits of both \( \varepsilon_{AB}^* \) and the correlation coefficient of items A and B. However, this modification reduces the direct relationship with the prediction error and affects the performance of IRS. The definition of Diagnosys, \( \gamma_{AB} = f_{\gamma_{AB}} / f_{\gamma AB} \) shows that the error frequency \( f_{\gamma AB} \) is divided by \( f_{\gamma AB} \), and this cause the relationship between \( \gamma_{AB} \) and the prediction error is reduced. From these observations, it is reasonable that OT has the best performance.

The performance of knowledge structure-based adaptive testing (KSAT) is affected by the correct guessing. Two methods, most difficult item method (MDI) and prerequisite item (PI) method, were proposed to deduce the possibilities of guessing. The experimental results show that both methods can improve the effect of correct guessing and have better performances than original methods.

Since OT has the best performance, it was selected to implement the KSAT system. The performance of the KSAT system shows that under the 93 percent prediction accuracy, the use of test items is 11.42. That is, on average, students need only complete one third of 35 items in the original paper-based exam when the KSAT system is applied. This result is close to that of the simulation study in experiment 1 and shows that the simulation process adopted in this study is valid and suitable.

From the above discussions, this study has three contributions. First, some evaluation methods for KSAT algorithms were applied to find the best adaptive testing algorithm among domain experts, OT, IRS, and Diagnosys, and OT-based KSAT algorithm has the best and stable performance. Second, two methods, the most difficult item method (MDI) and the prerequisite item (PI) method were proposed to improve the effect of correct guessing in KSAT algorithms. Finally, an OT-based adaptive testing system was developed and evaluated. Upon completion of the adaptive test, a diagnosis profile about the student’s state of learning or understanding was provided to do the subsequent actions, such as tailored instruction or remediation in applied educational settings. Another two directions are considered in the future study. First, OT, IRS, and Diagnosys were used to analyze the ordering relationship of dichotomous items, those methods for polytomous items will be considered in the next step. The second is to develop constructed response items and their automatic scoring mechanism to enhance the function of the KSAT system.

Acknowledgement

The authors would like to thank the National Science Council of the Republic of China and the National Taichung University of Education for financially supporting this research under contract numbers: NSC-92-2521-S-142-003, NSC 97-2511-S-142-004, NSC 98-2410-H-142-005-MY2, NSC-100-2410-H-656-007 and 98T202-3.

References


Exploring Learner Attitudes toward Web-based Recommendation Learning Service System for Interdisciplinary Applications

Hong-Ren Chen and Jhen-Gang Huang
Department of Digital Content and Technology, National Taichung University, Taichung 403, Taiwan // hrchen@mail.ntcu.edu.tw

(Submitted November 22, 2009; Revised January 26, 2011; Accepted March 31, 2011)

ABSTRACT

The booming digital-content industry has resulted in an increasing number of e-learning Internet websites that provide online learning services. Recommendations for learning sites are used by diverse learners to identify the most appropriate learning resources. However, research into recommendations about learning has concentrated primarily on suggestions for teaching materials to be used in a particular learning program, and studies have focused on improving the techniques underlying these recommendations without addressing the possibilities for interdisciplinary learning or changes in learner preferences. This study implemented and evaluated a system that recommends interdisciplinary pedagogical resources while recognizing changes in the foci of educational efforts and interests. This system was designed to satisfy individualized learning demands by recommending the most suitable content within an interactive context. Evaluations of learners’ attitudes and intentions after using the proposed system revealed widespread acceptance. Data analysis demonstrated that system stability and interface satisfaction were the key factors in learners’ attitudes. Indeed, interface satisfaction was the strongest predictor of recommendation accuracy and content satisfaction.

Keywords
Recommendation learning, Interdisciplinary application, Learner attitudes

Introduction

The huge quantity of learning resources currently available online has led to information overload, and traditional search engines can no longer meet the needs of all e-learners (Howe & Dreiling, 1997; Pinkerton, 2000; Yan & Garcia-Molina, 1995; Zamboni, 1998) because these tools often provide irrelevant information while ignoring related content (Goldberg, Nichols, Oki, & Terry, 1992). These so-called recommendation systems aim to identify the appropriate learning materials among the e-learning resources available to help learners make appropriate choices (Resnick & Varian, 1997). Although only high-quality materials are presumably recommended to learners, such materials may not meet learners’ expectations. For this reason, most current studies have focused on identifying information that meets the particular needs of learners (Wei, Moreau, & Jennings, 2005).

Current educational recommendation services use either explicit or implicit ratings. Calculations of explicit ratings are based on learners’ input on such issues as interest in courses, quality of learning units, and difficulty of teaching materials. Thus, although the accuracy of these recommendations is high, excess data may present a burden to learners. Lang (1995) proposed a recommendation service that rated articles via browsers that would analyze and reorganize the ratings of previous readers to offer further recommendations to learners who have not yet read the articles (Kruwich & Burkey, 1997). Unlike explicit ratings, which are based on learners’ expressed learning preferences, implicit ratings automatically record learning paths (e.g., learning materials read by learners, time spent on each learning unit, and frequency of daily visits to the learning website) for analysis. However, greater numbers of paths would require longer amounts of time for the calculation of recommendation rules; such implicit ratings are usually used to develop personalized recommendations for websites. A typical example of an implicit rating system was presented by Rucker and Polanco (1997), who suggested that the addition of a website to a list of bookmarks indicates interest in that website. Therefore, similar preferences can be calculated on the basis of the bookmarked sites and used to recommend additional websites.

Previous research on the recommendation of learning services has usually focused on recommendations for one course, with the goal of continuously creating different recommendation algorithms to enable learners to obtain more precise content (Johannes, Matthias, Christoph, & Ralf, 2008; Tasi, Chiu, Lee, & Wang, 2006; Wei et al., 2005). However, these studies have neglected immediate individual interactive learning models based on interdisciplinary learning, new fields of learning, and changes in academic interests. Content restricted to single subject cannot meet...
the diverse needs of modern industrial development, and interdisciplinary learning and cooperation have emerged as approaches for learners trying to solve complicated professional issues. From a psychological perspective, interaction between different fields and domains of knowledge can trigger learners’ intellectual potential, and exposure to multiple domains can enhance the scope, depth, and novelty of knowledge (Johansson, 2006). Thus, this study examined a platform for a recommendation service that supports interdisciplinary learning and used the advantages of explicit and implicit ratings to develop a mechanism for detecting changes in academic interests. Such a mechanism enabled the system to recognize each learner’s learning process and to actively recommend content from the specific fields required by learners. It also provided precise recommendations for learning content and reused pedagogical resources.

This study was designed to implement an interdisciplinary recommendation system that provides educational suggestions while monitoring changing academic interests to meet the individualized interactive learning demands of learners by recommending content that is appropriate to their current needs. We also evaluated learners’ attitudes and intentions with respect to using an interdisciplinary recommendation service. The study examined the predictive relationships among five factors: system stability, interface satisfaction, recommendation accuracy, satisfaction with content, and system acceptance. In this way, the functions of our system could be expanded and altered on the basis of the needs and attitudes of learners.

**Literature Review**

Social–cognitive theory (SCT) describes an individual’s behavior in terms of a reciprocal feedback system involving environment, personal attributes, and behaviors (Bandura, 1977). SCT has been applied in studies of education, decision making, management, and computer skills (Compeau & Higgins, 1995; Wood & Bandura, 1989; Zimmerman, 1990). The theory of reasoned action (TRA), as proposed by Fishbein and Ajzen (1975), can integrate previous theories on the influence of attitudes on behavior. This theory assumes that behavior can be controlled by an individual’s will, which could thus be used to predict and explain individual behaviors. However, the theory is usually restricted by many factors that significantly reduce its explanatory power regarding individual behavior. To enhance the predictive power of the TRA, Ajzen (1985) added factors related to perceived behavioral control over two dimensions that presumably influence behavioral intentions, namely attitudes toward behavior and subjective norms, to form the theory of planned behavior (TPB). The technology acceptance model (TAM), developed by Davis (1989) based on the TRA, provides general explanations about an individual’s actual and predicted use of information technology. The TAM has served as a theoretical foundation for investigations into the influences of external variables and users’ attitudes on use intentions. These studies have suggested that perceived usefulness and ease of use influence attitudes regarding the use of technology and further affect behavior. Previous empirical studies have found that users’ attitudes directly influence use intentions. At present, the TAM is widely applied in various fields such as information systems (Heijden, 2003), e-learning systems (Ndubisi, 2006), online shopping, and online gaming (Gefen, Karahanna, & Straub, 2003; Hsu & Lu, 2004).

Most research papers have focused on improving the techniques used in web-based recommendation learning systems (Chen, 2005; Cunningham & Frank, 1999; Krulwich & Burkey, 1997; Wang, Chuang, Hsu, & Keh, 2004). Nevertheless, understanding learners’ attitudes can also help to expand system functions and meet learners’ needs, which should further increase the impact of learning and enhance satisfaction with the learning process. The present study differs from previous studies in providing only a description of technology-oriented deduction (Chen, 2005; Cunningham & Frank, 1999; Krulwich & Burkey, 1997; Wang, Chuang, Hsu, & Keh, 2004). We used a questionnaire format to understand the reactions of learners to using a web-based learning system for recommending interdisciplinary learning resources. The attitudes of learners are crucial to the successful development of such a web-based recommendation learning system. For this reason, we adopted the Three-tier Use Model (3-TUM) proposed by Liaw (2007), which emphasizes individual attitudes toward information technology and has been widely employed in many important studies of user attitudes (Liaw & Huang, 2003; Liaw, Huang, & Chen, 2007; Lai, Huang, Liaw, & Huang, 2009). Perhaps most importantly, we integrated SCT, TPB, and TAM while simultaneously considering individual experiences, system quality, affective and cognitive factors, and behavioral intentions (Triandis, 1971; Davis, Bagozzi, & Warsaw, 1989; Moon & Kim, 2001; Liaw & Huang, 2003). We used a validated survey to quantitatively examine the influences of individual experiences and system quality on the affective and cognitive reactions of users; we were then able to ascertain how users’ reactions then affected behavioral intentions.
Previous studies have explored the factors that influence the use and acceptance of e-learning systems. Ozok, Fan, and Norcio (2010) reported a significant positive correlation between a system’s predictive consistency and the accuracy of its recommendations and noted that learners considered system reliability to be an important factor in their use of a system. Interface satisfaction refers to the convenience experienced by learners when using a system insofar as it conforms to the operational habits of learners (Nielsen, 1993; Rushieks & Rushineks, 1986). Chien (2009) noted a significant positive correlation between interface satisfaction and content satisfaction. E-learning systems must offer stability and user-friendly interface to continue to attract learners, motivate them to visit learning websites, and explore the relationships among predictive factors and systems-management issues in the affective and cognitive reactions of learners (Shneiderman, 1997). Literature using specific reading or instructional materials as indicators of use intentions (Chen & Yeh, 2008; Hu & Pu, 2009; and Kuan, 2004) has found a significant positive correlation between recommendation accuracy and intent to use e-learning resources. Recommendation accuracy refers to whether the suggested content satisfies learners’ demands and influences their future use intentions. A study conducted by Chang (2010) on the use and acceptance of mobile learning communities found that satisfaction with the educational content affected acceptance of the system.

Thus, the accuracy of the recommendations and the satisfaction with the content offered by e-learning systems can be investigated as measurable variables (Chang, 2010). Indeed, Baker-Eveleth, Eveleth, O’Neill, and Stone (2006) examined laptop exams; Cheng-Chang, Gunter, Sivo, and Cornell (2004–2005) explored e-learning management systems; and Ndubisi (2006) analyzed the online learning community’s acceptance of a particular system. According to research performed by Simonson, Smaldino, Albright, and Zvacek (2000), the following four important factors enhanced learners’ use of information technology: learners’ attitudes, experiences, cognitions, and learning styles. Of these factors, learners’ attitudes were found to be the most important indicator of satisfaction. After gaining learners’ acceptance, a learning system should meet learners’ demands by increasing its functions to enhance learning.

Hypothesis 1: Five factors affect use intention with web-based systems designed to recommend interdisciplinary educational resources to learners: system stability, interface satisfaction, recommendation accuracy, content satisfaction, and system acceptance.

The 3-TUM research framework proposed by Liaw (2007) is currently widely used in many important studies on learners’ attitudes (Lai et al., 2009; Liaw & Huang, 2003; Liaw et al., 2007). The major conceptual foundation of this approach draws on SCT, TPB, and TAM to explore the attitudes of the users of information technology at three different levels: individual experiences and system quality, affective and cognitive reactions, and behaviors and intentions. The level of individual experiences and system quality identifies the events that can influence affective and cognitive reactions. Quantification at the affective and cognitive level allows for predictions about intentions and behaviors (David et al., 1989; Liaw, Chen, & Huang, 2008; Triandis, 1971). In other words, individual experiences and system quality can positively influence affective and cognitive experiences. At the same time, affective and cognitive dimension can also positively impact behaviors and intentions. Extant literature shows that system stability and interface satisfaction were included in the category of system quality, whereas recommendation accuracy and satisfaction with content constituted cognitive factors, and behaviors and intentions were related to system acceptance.

Hypothesis 2a: System stability and interface satisfaction can positively predict recommendation accuracy and content satisfaction.

Hypothesis 2b: Recommendation accuracy and content satisfaction can positively predict system acceptance.

Development of a web-based learning system for recommending interdisciplinary learning resources

During the process of e-learning, learning guides help each learner navigate through materials according to their abilities, interests, habits, and needs. Traditional classroom teaching uses the same schedule for all learners under the assumption that all have similar capabilities. Learners using recommendation services are regarded as unique individuals with particular learning abilities and preferences. They are allowed to use materials at their own pace and are expected to have personalized learning experiences. Moreover, the system records every completed learning process to enable understanding of the learning schedule and status of each learner as well as to provide learners with clear insight into their own status as learners.
The proposed recommendation service is presented in Figure 1. Learning objects are transformed into vector-based representations through a learning-object extractor and are then stored in the learning-object domain. The learning-content-management subsystem (LCMS) stores the learning objects in the database in the form of a shared content-object reference model (SCORM). Learning objects can be in the form of text, graphics, videos, and simulation clips, which enable their repeated use as the system expands to include additional learners. Initially, new learners will be guided to a survey to examine their interests, and these data will be used to establish individualized learning profiles. Previous users will be sent directly to the learner-profile domain to obtain their personal profiles. Next, the system’s recommendation operations will identify the appropriate learning objects. Learning objects with the highest scores will be recommended to learners, as shown in Figure 2.

The following four courses will be saved in the database of the system: basic computer concepts (BCC), databases, networks, and web design. This study classified these courses according to the characteristics of their teaching materials. The difficulty of materials was divided into low, medium, and high. The time required to learn the materials was classified as within 30 minutes, 30–60 minutes, and more than 60 minutes. The types of teaching materials included introductory, theoretical, practical, and mathematical. Introductory material presents foundational concepts, theoretical material teaches logic or epistemology, practical material involves the use of software, and mathematical material involves methods of calculation. The extended object-oriented portal system (XOOPS) was selected as the infrastructure for this web-based service.

The recommendation function is key success factor to recommend learning service system which mainly introduces the advantages of explicit and implicit ratings to proceed a design. A cross validation is conducted on the data.
generated from explicit and implicit ratings, and add on the hottest recommended principle Top-N into the attribute value of learning interest table and proceed sieving (Karypis, 2001; Sarwar, Karypis, konstan, & Riedl, 2000), to sieve out the course materials that interests no learners, and to use association rules at the same time to analyze the e-portfolio of each learner to achieve the learning contents required by the learner to enhance the accuracy of recommended course materials (Agrawal & Srikant, 1994; Chen & Liu, 2004). In consideration of interdisciplinary learning of learners, we design an effective multidimensional system to detect the change in learning; including three following operating rules: (1) if learners choose learning contents of different subjects from learning interest table, and when the number of clicks exceeds minimum support count, the system will take the new-added field and will recommend the learning materials of this new-added subject automatically. (2) if learners did not click on the learning contents recommended by the system when the number of clicks exceeds minimum support count, the system will acknowledge a change of learning interest and remove the recommended course material of this subject. (3) if the system reports a failure on recommendation, in other words, did not conform to the learning contents that learners want, then a weighed adjustment strategy on explicit and implicit ratings will be enabled to increase the weighted of e-portfolio analytic results and to proceed a recommendation to learning contents again.

For example, learners only click on the e-learning courses while setting their learning interest, after a period of time, perhaps, they will see a need in interdisciplinary learning, as implementation of e-learning website would require homepage design knowledge, the system will automatically detect number of clicks on learning contents of webpage design. In order to avoid error detection of learners caused by random browsing, number of clicks has to exceed minimum support count defined by the association rule (Agrawal, Imieliński, & Swami, 1993) to allow the system to acknowledge the new-added field that learners belong to, and automatically recommends the learning contents of webpage design course to learners.

After the system recognizes these correct clicks, new lists are expanded and new recommendation contents can be formed. The recommendation function is shown as Figure 3 and the following are recommendation measures by an example:

- Step 1: selection of interested learning items in the system and setup of learning profile.
- Step 2: records of interested learning items. Filtering learning contents in accordance with the learning profile orderly based on interested subject, level of learning content, learning time and learning style.
- Step 3: analysis and cross validation over the path linked list formed through the association rules, in order to find out a learning content with functions over minimum support count.
- Step 4: detection of changing learning process to adjust setting of learning fields.
- Step 5: a customized recommended learning list is produced by calculation of recommendation functions based on the weighted ratios. Namely, the system demonstrates the recommended learning contents based on the ratio of implicit and explicating rating setting by the system administrator.
- Step 6: readjust the recommended learning lists once the recommendation fails by starting weighted adjusting strategy.

Figure 3 presents the illustration operations of a recommendation function. In the first step, learners set a learning profile and subject of BCC. Learning difficulty is the primary level, the learning time is 30-60 minutes, and learning type is practice. In this step, learners access setting in registration of the recommendation learning system. The system will automatically record each learner in a database. In the second step, the system will show a screen database of learning materials according to each learner’s learning profile. These two steps are explicit rating. The third step is implicit rating. The system will record each learner’s teaching materials as a learning path and select teaching materials upon association rules. For instance, the P1 learning path is shown by \{B1, B2, B3\}, and it means the order of learners’ reading materials is B1, B2, and B3. Therefore, learning paths P1, P3, and P4 would be selected with the minimum support count assumed as 3. In the fourth step, the system will calculate the mixture of explicit rating and implicit rating, and include the teaching materials recommended learning list. After detection of a changing learning process, the mechanism demonstrates that although the web design course is not the subject listed as interesting learners in learning profile, the learners usually read the teaching materials for web design course, which is not random reading. The system will recommend teaching materials regarding the web design course, as it is a mechanism of interdisciplinary recommendation learning. In the fifth step, a recommended learning material list will be shown to registered learners. However, if learners do not read the recommendation teaching materials, the recommendation will fail. The system will continue with the next recommendation by weighted adjusted strategy to offer more accurate recommended teaching materials.
Research Design

Our sample consisted of 195 students in IT-related departments at universities located in central Taiwan. A questionnaire was distributed to these students 6 weeks after they had used the web-based interdisciplinary recommendation system. Participants were engaged in using this system for an average of 72 hours prior to completing the survey. We received 182 valid questionnaires, which represented a response rate of 93.33%. The purpose of this experiment was to investigate learners' attitudes about the use of an interdisciplinary recommendation system. The items contained in the scale measuring these attitudes were adapted from the extant literature on web-based learning. Well-known experts were invited to discuss the validity of the questionnaire based on the relationship between our research purpose and the content and phrasing of the items in the scale. Questions with poor discriminative validity were eliminated, and ambiguous phrases were modified. A total of 22 questions using a 7-point Likert scale from 1 to 7 (representing strongly disagree to strongly agree) were included in the final instrument.

This study based on theories presented in previous studies (Chen, 2007; Liaw et al., 2008; Moon & Kim, 2001; Vankatesh, 1999), used a questionnaire methodology to understand the reactions of learners to a web-based system for recommending educational resources because learners' attitudes are crucial to the success of any web-based recommendation system. On the basis of our literature review, this study focused on system stability, interface satisfaction, recommendation accuracy, satisfaction with content, and system acceptance as key contributors to the attitudes of learners toward such a system.

Data Analysis and Results

Prior to performing factor analysis, we used the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity to examine the correlations among variables to determine their suitability for factor analysis. The test results revealed a
KMO value of 0.83, suggesting close correlations among variables. The chi-square value for Barlett's test of sphericity was 2798.70, which was significant and indicative of common factors that rendered the data appropriate for follow-up factor analysis. The study used principle-component analysis (PCA) for conducting the factor analysis in two steps. The first step, factor extraction, was used to calculate the variance shared among all variables. Eigenvalues higher than 1 were identified following Kaiser’s rule, which defines the factors that can be used to explain variance in the results (Kaiser, 1958). We also used a scree test to identify the factors selected during extraction (Cattell, 1966). The second step, factor rotation, was performed to enhance our ability to explain factor loading. Factors with eigenvalues higher than 1 were selected for factor rotation. Varimax orthogonal rotation was used in this analysis because the simple structure of this approach easily explained the results.

Table 1 shows the five factors extracted in the factor analysis using the minimum factor-loading standard of 0.6 suggested by Hair, Black, Babin, Anderson, and Tatham (2009). The eigenvalues and percentages of variance obtained using Varimax rotation are also presented. Five items were contained in factor 1, five in factor 2, four in factor 3, four in factor 4, and four in factor 5. The five factors were system stability, interface satisfaction, recommendation accuracy, system acceptance, and content satisfaction. Table 2 shows the items included in the factors, the factor loadings, and their levels of reliability.

Table 1. Total variance explained and percentage of variance by Varimax method

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Percentage of variance</th>
<th>Cumulative percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.68</td>
<td>16.73</td>
<td>16.73</td>
</tr>
<tr>
<td>2</td>
<td>3.39</td>
<td>15.43</td>
<td>32.16</td>
</tr>
<tr>
<td>3</td>
<td>3.14</td>
<td>14.25</td>
<td>46.41</td>
</tr>
<tr>
<td>4</td>
<td>2.97</td>
<td>13.51</td>
<td>59.92</td>
</tr>
<tr>
<td>5</td>
<td>2.78</td>
<td>12.64</td>
<td>72.56</td>
</tr>
</tbody>
</table>

We calculated Pearson’s correlation coefficients to further assess the predictive power of the variables included in this study (Stevens, 1992). Path analysis, a statistical method for elucidating relationships among multiple variables, uses regression analysis to understand the relative influences of variables. However, problems of multicollinearity must be eliminated. Following the suggestions offered by Blumberg, Cooper, and Schindler (2008) and Neter, Wasserman, and Kutner (1990), we ensured that (1) the correlation between variables was less than 0.8 and (2) the variance inflation factors (VIFs) were lower than 10. Table 3 shows the Pearson’s correlation coefficients between variables; significant correlations are those with coefficients less than 0.8. The results of stepwise multiple regression analyses are shown in Table 4, which also shows that the VIFs were lower than 10. The results of our research model are also shown in Figure 4. The first regression analysis examined the effect of system stability and interface satisfaction on recommendation accuracy. The results suggest that these two factors were independent predictors of recommendation accuracy ($F(2, 179) = 37.24, p = 0.000 <0.001, R^2 = 0.30$). The second regression analysis examined the predictive power of satisfaction with content on system stability and interface satisfaction. Simply stated, the results indicated that system stability and interface satisfaction predicted content satisfaction ($F(2, 179) = 43.23, p = 0.000 <0.001, R^2 = 0.33$). The final regression analysis examined the predictive power of recommendation accuracy and content satisfaction on system acceptance. The results showed that recommendation accuracy was more strongly predictive of system acceptance than was satisfaction with content ($F(2, 179) = 34.11, p = 0.000 <0.001, R^2 = 0.28$).

Table 2. Reliability and factor analysis

<table>
<thead>
<tr>
<th>No.</th>
<th>Items</th>
<th>Factor loading</th>
<th>Cronbach’s α</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The speed of system connection is good and no need to waste much time in waiting.</td>
<td>0.871</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The browse of teaching materials won't cause computers to crash.</td>
<td>0.771</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>I satisfied with the system connection speed.</td>
<td>0.805</td>
<td>0.88</td>
</tr>
<tr>
<td>4</td>
<td>There is no response delay on the system.</td>
<td>0.755</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>As a whole, I am satisfied with the system stability</td>
<td>0.726</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>System menu and links are clearly marked, which makes it easy and comfortable for me to browse.</td>
<td>0.740</td>
<td>0.87</td>
</tr>
<tr>
<td>7</td>
<td>I feel the layout and the color arrangement of the system interface are</td>
<td>0.856</td>
<td></td>
</tr>
</tbody>
</table>
8. I feel the layout and the color arrangement of the system interface make it easy to use.

9. I feel the steps of system setting are easy and convenient.

10. Overall, I am satisfied with the system's operating interface.

Factor 3: Recommendation accuracy

11. When my learning preferences change, I find the teaching materials recommended by the system suit my needs.

12. In the interdisciplinary learning, I find the teaching materials recommended by the system suit my needs.

13. The teaching materials recommended by system's recommendation function suit my needs.

14. All in all, I am satisfied with the recommendation accuracy of the system.

Factor 4: System acceptance

19. I believe the system is helpful to my learning.

20. I will use the system to assist my learning.

21. I will use the system to learn in the future.

22. I think it is necessary to use the system in learning.

Factor 5: Content satisfaction

15. The contents of the teaching materials recommended by the system suit my learning preferences.

16. In terms of learning time, I am satisfied with the teaching materials recommended by the system.

17. In terms of learning styles, I am satisfied with the teaching materials recommended by the system.

18. As a whole, I am satisfied with the teaching materials recommended by the system.

The results of the data analysis showed that system stability and interface satisfaction were the most important contributors to the attitudes of learners. System stability and interface satisfaction predicted recommendation accuracy and content satisfaction, respectively. Recommendation accuracy and content satisfaction also predicted system acceptance. These results are consistent with those reported by Ozok et al. (2010), Chien (2009), Hu and Pu (2009), and Chang (2010). Interface satisfaction was found to be the best predictor of recommendation accuracy and content satisfaction. After learners accept recommendation systems for educational resources, the accuracy of recommendations has the strongest predictive power with respect to learner reactions.

Table 3. Correlation analysis

<table>
<thead>
<tr>
<th>variables</th>
<th>System stability</th>
<th>Interface satisfaction</th>
<th>Recommendation accuracy</th>
<th>System acceptance</th>
<th>Content satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>System stability</td>
<td>1</td>
<td>0.467**</td>
<td>0.500**</td>
<td>0.509**</td>
<td>0.471**</td>
</tr>
<tr>
<td>Interface satisfaction</td>
<td>1</td>
<td>0.589**</td>
<td>0.500**</td>
<td>0.589**</td>
<td></td>
</tr>
<tr>
<td>Recommendation accuracy</td>
<td>1</td>
<td>0.585**</td>
<td>0.480**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System acceptance</td>
<td>1</td>
<td></td>
<td></td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Content satisfaction</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p<0.01

Table 4. Regression results of predicted path relationships

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Dependent variable</th>
<th>Independent variables</th>
<th>β</th>
<th>R² change</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2a</td>
<td>Recommendation accuracy</td>
<td>System stability</td>
<td>0.36</td>
<td>0.18</td>
<td>0.000</td>
</tr>
<tr>
<td>H2a</td>
<td>Recommendation accuracy</td>
<td>Interface satisfaction</td>
<td>0.35</td>
<td>0.12</td>
<td>0.000</td>
</tr>
<tr>
<td>H2a</td>
<td>Content satisfaction</td>
<td>Interface satisfaction</td>
<td>0.52</td>
<td>0.30</td>
<td>0.000</td>
</tr>
<tr>
<td>H2a</td>
<td>Content satisfaction</td>
<td>System stability</td>
<td>0.17</td>
<td>0.03</td>
<td>0.008</td>
</tr>
<tr>
<td>H2b</td>
<td>System acceptance</td>
<td>Recommendation accuracy</td>
<td>0.42</td>
<td>0.25</td>
<td>0.000</td>
</tr>
<tr>
<td>H2b</td>
<td>System acceptance</td>
<td>Content satisfaction</td>
<td>0.18</td>
<td>0.03</td>
<td>0.015</td>
</tr>
</tbody>
</table>
Discussion

The results obtained from the data analysis presented in Tables 1 and 2 are consistent with previous results related to learning attitudes (Liaw & Huang, 2003; Liaw et al., 2008), and they also support Hypothesis 1 of this research. In terms of the services provided by web-based recommendation services, learner attitudes can be divided into five factors: system stability, interface satisfaction, recommendation accuracy, satisfaction with content, and system acceptance. The factor loading for each factor was greater than 0.6, with Cronbach’s $\alpha$ exceeding 0.8, which demonstrated the distinguishability and reliability of the factors and is consistent with the proposals of Hair et al. (2009). Two major factors should be considered in the development of a web-based recommendation environment: system stability and interface satisfaction. Delone and McLean (1992) advocated system stability as the key factor in ensuring the quality of an information system, whereas Liaw et al. (2008) found that learners were more concerned about user interface. On the other hand, the present study found that positive evaluations of usefulness and positive behavioral intentions were related to the accuracy of recommendations and satisfaction with content.

The statistical results of the regression analysis presented in Table 4 quantify learner attitudes toward services provided by the web-based recommendation learning system, validate Hypotheses 2a and 2b of this research, and support the 3-TUM model. When system stability and interface satisfaction were used to predict recommendation accuracy, system stability explained 18% of the variance in recommendation accuracy; when system stability and interface satisfaction were used to predict content satisfaction, interface satisfaction explained 30% of the variance in content satisfaction; and when recommendation accuracy and content satisfaction were used to predict system acceptance, recommendation accuracy explained 25% of the variance in system acceptance. Thus, interface satisfaction was the strongest predictor of recommendation accuracy and content satisfaction, whereas acceptance of a system offering web-based recommendations for learning resources was the most significant predictor of recommendation accuracy.

Conclusion

The ongoing development of network technologies has resulted in the availability of digitized knowledge on many digital-learning websites. Learners often get lost when faced with huge quantities of online learning resources, and consequently, many learning services focus on providing recommendations for course materials based on learners’ interests. Learners no longer accept the need to engage in effortful searches for suitable digital teaching materials or for knowledge in other fields. Most previous studies related to recommendations for learning have focused on the accuracy of the digital teaching materials recommended for a single subject area but have neglected interdisciplinary learning, novel learning areas, and multidimensional recommendations. Our concept of a service offering interdisciplinary recommendations of learning resources would draw on massive online resources to identify precise interdisciplinary digital learning content for learners. This paper differs from past discussions of the technology-oriented basis of online environments designed to offer recommendations related to learning. Previous studies have focused on improving the performance of recommendation algorithms, whereas the present study focused on understanding the attitudes of learners and expanding the system functions to enhance learning effectiveness and satisfaction. Learner attitudes were surveyed after participants used the interdisciplinary recommendation learning system, and five factors were extracted: system stability, interface satisfaction, recommendation accuracy, content

Discussion

The results obtained from the data analysis presented in Tables 1 and 2 are consistent with previous results related to learning attitudes (Liaw & Huang, 2003; Liaw et al., 2008), and they also support Hypothesis 1 of this research. In terms of the services provided by web-based recommendation services, learner attitudes can be divided into five factors: system stability, interface satisfaction, recommendation accuracy, satisfaction with content, and system acceptance. The factor loading for each factor was greater than 0.6, with Cronbach’s $\alpha$ exceeding 0.8, which demonstrated the distinguishability and reliability of the factors and is consistent with the proposals of Hair et al. (2009). Two major factors should be considered in the development of a web-based recommendation environment: system stability and interface satisfaction. Delone and McLean (1992) advocated system stability as the key factor in ensuring the quality of an information system, whereas Liaw et al. (2008) found that learners were more concerned about user interface. On the other hand, the present study found that positive evaluations of usefulness and positive behavioral intentions were related to the accuracy of recommendations and satisfaction with content.

The statistical results of the regression analysis presented in Table 4 quantify learner attitudes toward services provided by the web-based recommendation learning system, validate Hypotheses 2a and 2b of this research, and support the 3-TUM model. When system stability and interface satisfaction were used to predict recommendation accuracy, system stability explained 18% of the variance in recommendation accuracy; when system stability and interface satisfaction were used to predict content satisfaction, interface satisfaction explained 30% of the variance in content satisfaction; and when recommendation accuracy and content satisfaction were used to predict system acceptance, recommendation accuracy explained 25% of the variance in system acceptance. Thus, interface satisfaction was the strongest predictor of recommendation accuracy and content satisfaction, whereas acceptance of a system offering web-based recommendations for learning resources was the most significant predictor of recommendation accuracy.

Conclusion

The ongoing development of network technologies has resulted in the availability of digitized knowledge on many digital-learning websites. Learners often get lost when faced with huge quantities of online learning resources, and consequently, many learning services focus on providing recommendations for course materials based on learners’ interests. Learners no longer accept the need to engage in effortful searches for suitable digital teaching materials or for knowledge in other fields. Most previous studies related to recommendations for learning have focused on the accuracy of the digital teaching materials recommended for a single subject area but have neglected interdisciplinary learning, novel learning areas, and multidimensional recommendations. Our concept of a service offering interdisciplinary recommendations of learning resources would draw on massive online resources to identify precise interdisciplinary digital learning content for learners. This paper differs from past discussions of the technology-oriented basis of online environments designed to offer recommendations related to learning. Previous studies have focused on improving the performance of recommendation algorithms, whereas the present study focused on understanding the attitudes of learners and expanding the system functions to enhance learning effectiveness and satisfaction. Learner attitudes were surveyed after participants used the interdisciplinary recommendation learning system, and five factors were extracted: system stability, interface satisfaction, recommendation accuracy, content
satisfaction, and system acceptance.

Working from an academic perspective, we proposed and implemented a service able to use innovative strategies to offer interdisciplinary recommendations and to automatically detect areas in which the interests and/or needs of learners have changed. This system can recommend suitable learning materials in the service of individualized learning. This study is also the first to use factor analysis in its examination of recommendation services for learning resources and to explore the attitudes and system acceptance of learners. Thus, our results support the 3-TUM model and the research hypotheses and serve as a foundation for the development of systems that recommend learning options. Working from a practical perspective, this study proposed four important elements of the proposed system: system stability, interface satisfaction, recommendation accuracy, and content satisfaction. Learners were most concerned with interface design and the accuracy of the recommendations for instructional material. Our data emphasized the importance of designing recommendation systems that allow learners to simply and conveniently use the services and that provide learning materials that conform to the learners' needs and interests. This study offers important contributions to the relevant literature by presenting innovative concepts regarding web-based interdisciplinary recommendations and by emphasizing the need for a system that is consistent with contemporary pedagogical conditions in its ability to enhance and effectively respond to diversity in learning.

Acknowledgements

This study is partially supported by the National Science Council of Taiwan under grants NSC97-2511-S-142-003 and NSC98-2511-142-005.

References


Passively Classifying Student Mood and Performance within Intelligent Tutors

Robert A. Sottilare¹ and Michael Proctor²

¹U.S. Army Research Laboratory – Human Research and Engineering Directorate, Orlando, Florida 32826, USA // ²University of Central Florida College, Department of Industrial Engineering and Management Systems (IEMS), Orlando, Florida, 32816, USA // robert.sottilare@us.army.mil // mproctor@mail.ucf.edu

(Submitted January 22, 2010; Revised April 2, 2011; Accepted May 20, 2011)

ABSTRACT

It has been long recognized that successful human tutors are capable of adapting instruction to mitigate barriers (e.g., withdrawal or frustration) to learning during the one-to-one tutoring process. A significant part of the success of human tutors is based on their perception of student affect (e.g., mood or emotions). To at least match the capabilities of human tutors, computer-based intelligent tutoring system (ITS) will need to “perceive” student affect and improve performance by selecting more effective instructional strategies (e.g., feedback). To date, ITS have fallen short in realizing this capability. Much of the existing research models the emotions of virtual characters rather than assessing the affective state of the student. Our goal was to determine the context and importance of student mood in an adaptable ITS model. To enhance our existing model, we evaluated procedural reasoning systems used in virtual characters, and reviewed behavioral and physiological sensing methods and predictive models of affect. Our experiment focused on passive capture of behaviors (e.g., mouse movement) during training to predict the student’s mood. The idea of mood as a constant during training and predictors of performance are also discussed.

Keywords

Adaptive tutoring system, Mood, Affect, One-to-one tutoring, Passive measures

Introduction

The goal of this research was to develop an adaptable ITS conceptual model that includes appropriate inputs to determine the affective state of the student being tutored. In support of this goal, we surveyed methods to allow an ITS to classify the affective state (e.g., mood or emotional state) of the student through the passive sensing and interpretation of correlated student behaviors and their physiological responses during training. We investigated methods of classifying affect for virtual characters through procedural reasoning systems and adapted these methods to human students to classify the student’s affective state. Understanding the student’s affective state, the ITS can use that information along with other student data (e.g., knowledge and progress toward goals) to select instructional strategies (e.g., direction) to optimize learning and performance.

The research discussed below offers perspectives on: one-to-one tutoring; affect and learning; the need for ITS to be capable of “perceiving” student affect; the design limitations of current ITS; models of affect; and enhanced ITS models. The main body of our research contains two primary objectives: the review of methods to sense student behaviors unobtrusively (passively) so as avoid any negative impact on the learning process; and the interpretation of sensed behaviors to build a predictive model of student affect as a basis for making decisions on the delivery of instruction.

A perspective on one-to-one tutoring

An ongoing goal in the research and development of ITS has been to increase their adaptability to better serve student needs (Heylen, Nijholt, op den Akker & Vissers, 2003; Hernandez, Noguez, Sucar & Arroyo-Figueroa 2006 and Sottilare, 2009). “The basic tenet of intelligent tutors is that information about the user (e.g., knowledge, skill level, personality traits, mood level or motivational level) can be used to modify the presentation of information so that learning proceeds more efficiently.” (Johnson & Taatgen, 2005, p. 24).

ITS offer the advantage of one-to-one tutoring where instruction is delivered at a tailored pace based on competency and progress toward instructional goals. The value of expert, one-to-one, human tutoring vice group tutoring (i.e. traditional classroom teaching) has been documented among students who often score 2.0 standard deviations higher
than students in a conventional classroom (Bloom, 1984). Other advantages of one-to-one, human-tutoring over classroom settings is that students ask approximately 26.0 questions per hour versus less than 0.20 questions per hour in the classroom setting (Dillon, 1988; Graesser & Person, 1994). This higher rate of interaction provides additional learning opportunities for weaker students. Stronger students ask fewer questions, but these questions tend to be “deep-reasoning questions” (Person & Graesser, 2003).

Loftin, Mastaglio & Kenney (2004) assert that while one-to-one human tutoring is still superior to ITS in general, the one-to-one human tutoring approach is neither efficient nor cost-effective for training large, geographically-distributed populations (e.g., military organizations or large multi-national corporations). Given large and potentially diverse student populations, one goal of our research was to identify methods to improve the adaptability of ITS to support one-to-one tutoring. We explored methods to increase ITS “perception” of the student’s affective state and thereby theoretically increase the potential of the ITS to effectively adapt to a given student during one-to-one, computer-based instruction.

A perspective on affect and learning

Linnenbrink and Pintrich (2002) identified a connection between affect and learning. They found that many students experience some confusion when confronted with information that does not fit their current knowledge base, but those in a generally positive affective state adapted their known concepts to assimilate the new information. Students in a generally negative state usually reject this new information. This infers the need for tutors (human or otherwise) to be able to perceive and address the affective state of the students when formulating instructional strategies.

Fundamental to our research is an understanding of affect and how it might be modeled within ITS. Affect includes personality traits, mood and emotions which vary in duration, influence and cause (Gebhard, 2005). Affect is evident through student behaviors and physiological changes and may also be measured via self-report surveys. Each method has limitations. Self-report measures may be biased by the student’s desire to conform to expectations and may not accurately reflect their true affective state. Physiological measures can vary from person to person and may be misinterpreted. Hybrid approaches that use combinations of these measures to confirm affect are considered more reliable. Being able to accurately predict affect is the first step in using it to select appropriate instruction.

An ITS should choose the content and the method of instruction based on both the student’s competence and affective state (e.g., emotions) in the same way “an experienced human tutor manages the emotional state of the student to motivate him and to improve the learning process” (Hernandez et al., 2006). Rodrigues, Novais & Santos (2005) affirmed that an ITS “must be capable of dynamically adapting and monitoring each student”. This highlights the need for computer-based tutors to be able to perceive student state and use this information in formulating instruction.

Design limitations of ITS

Affect has been modeled extensively within ITS (Core et al., 2006; Heylen et al., 2003; Graesser et al., 2001), but have generally been used to represent the tutor’s affect rather than perceiving affect in students. Incorporating affect perception within ITS is recognized as a key element in the learning process (Bickmore & Picard, 2004; Burleson & Picard, 2004; Johnson, Richel & Lester, 2000; Picard et al., 2004), but Picard (2006) notes the design limitations of ITS including their inability to: recognize student affect; respond appropriately to the student affect; and appropriately express emotion. Picard suggests that the “simplest set” of emotions for an ITS to recognize are the emotions each of us is born with: pleasure, boredom and frustration. The ability to sense and predict even this small set of emotions would greatly expand the capability of ITS to adapt instruction to the needs of the student and optimize learning and performance.

Further expanding the need for ITS to perceive student affect and adapt instruction, Alexander, Sarrafzadeh and Fan, (2003) argued that “an important factor in the success of human one-to-one tutoring is the tutor’s ability to identify and respond to affective cues given by the student”. We also expected that ITS that model affect would be more effective than equivalent ITS that do not model affect.
Picard (2006) asserted that “no matter how intelligent a tutor is, it will eventually become annoying if it does not have emotional intelligence”. Emotional intelligence is defined as “a set of skills hypothesized to contribute to the accurate appraisal and expression of emotion in oneself and in others, the effective regulation of emotion in self and others, and the use of feelings to motivate, plan, and achieve in one’s life” (Salovey & Mayer, 1990). In other words, ITS that have emotional intelligence are capable of: recognizing the student affect; making the student aware so he can participate in managing his affective state; providing options (e.g., strategies) for the student to control his affective state; using emotion to motivate the student to achieve established objectives.

Today’s ITS provide limited adaptive training capabilities. Sessink et al. (2007) assert that “most traditional learning material [for ITS] targets the ‘average student’, and is suboptimal for students who lack certain prior knowledge, or students who have already attained some of the course objectives”. Csikszentmihalyi (1990) stated that ITS should match “adequate challenge with skill in service of Flow, or optimal experience” within the learning system. Failure to provide adequate challenge could result in boredom or frustration. It is desirable for the ITS to know everything possible about the student’s affect to craft a tailored learning experience.

**ITS models**

We chose to begin with a documented standard ITS model. Beck, Stern and Haugsja’a’s (1996) Intelligent Tutoring System Model (shown in Figure 1) illustrates the interactions ITS components including: domain knowledge which provides basis for course content, a student model which generally contains information about the progress of the student relative to learning objectives; an expert model which defines the performance of an ideal student; a pedagogical module which makes decisions on instruction and feedback to the student; and finally, a communication interface which presents instruction and feedback to the student. Given the goal to expand the student model to include affect and optimize decisions on instruction, Beck’s ITS model has some significant limitations including the inability to sense the student’s behavior and physiological state.

![Figure 1. Beck, Stern and Haugsja’a’s (1996) Intelligent Tutoring System Model](image)

Based on the current limitations of ITS and the emphasis in the literature on understanding the affective state of the student in one-to-one tutoring situations, Sottilare (2009) adapted Beck, Stern and Haugsja’a’s (1996) ITS model as
shown in Figure 2 to include the student’s affective state and methods to assess affect (e.g., behavioral, physiological measures). This model enhanced Beck, Stern and Haujsjaa’s model by adding capabilities to sense behaviors and physiological data, and transfer this data to the student model to aid in assessing the student’s affective state; and by adding a real-time flow of affective state data from the student model to the pedagogical model so the student’s affective state can play a part in instructional strategy decisions. Sottilare’s (2009) model was used as the foundation for integrating affective classification models based on procedural reasoning systems as discussed below.

Procedural reasoning systems

Like the logic within ITS, procedural reasoning system (PRS) for virtual humans act as intelligent agents which manage processes (e.g., sensing processes). For this reason, we decided to evaluate the sensing and reasoning processes of PRS to see if there were any transferrable methods for use in ITS. We evaluated several existing models used in embodied characters and their evolution in providing more complex interactions within intelligent agents. We also sought to reuse key elements and adapt these models to support a comprehensive ITS model that includes affect.

First, we considered the basic procedural reasoning system (PRS) model developed by Georgeff and Lansky (1987). This model is based on a belief-desire-intention framework as illustrated in Figure 3 and includes the following states and processes: beliefs state are facts about the world which are perceived to be true; goals/desires state are a set of needs/wants to be realized; knowledge areas are declarative procedure specifications or sequences; interpreter process is the analysis of relevant goals/desires, beliefs and a set of relevant knowledge areas to determine intentions; intentions state is a set of planned actions; action process include actions taken in order to achieve a goal or satisfy a desire; environment state include any external influencing factors; and finally, the perception process which senses changes in the environment after actions take place.

Parunak et al., 2006 enhanced Georgeff and Lansky’s model with the addition of OCC model of emotions (Ortony, Clore & Collins, 1988). In this enhanced model, beliefs feed an appraisal process as well as the interpreter. The appraisal process determines which of the 22 emotions defined in the OCC model to output to the perception and interpreter processes. Parunak et al., (2006) expanded the PRS model again with the addition of disposition, “a parameter that distinguishes the varying susceptibility of different agents to various emotions”. Parunak considered disposition to be constant over the time horizon of the simulations that they used for experimentation. For our purposes, we considered disposition in terms of a student’s varying susceptibility to various emotions. In this light, disposition might be influenced by factors like personality, mood or competency in the learning domain in which the student is engaged.

As noted previously, Linnenbrink and Pintrich (2002) identified a connection between affect and learning. For instance, a personality factors like openness might influence a student’s disposition to learning new information. Competency might influence confidence and thereby dispose a student toward specific emotions like frustration if the instruction presented is not in line with their competency level. Therefore, an enhanced student model within an ITS that includes affect along with passive, real-time sensory inputs (e.g., behavioral and physiological measures)
would make that ITS better able to adapt to a student’s needs. In addition to its effect on perception, note that changes in emotion might also positively or negatively influence desires and plans in this enhanced student model.

Default values for personality, mood and competency are based on instruments like the Big Five Personality Test, the Self-Assessment Manikin (SAM) for mood (Lang, 1980) and a skills/knowledge pre-test to determine initial competency. Emotions are assessed probabilistically by frequent appraisals of behaviors and physiological measures. Competency might also be assessed periodically to determine if the instruction is having the desired effect.

Finally, integrating these concepts with the Intelligent Tutoring/Coaching Model in Figure 2, we see the potential for a more adaptive ITS model in Figure 4 that accounts for affect, competency, goals/desires, plans and beliefs to tailor instruction to each student’s needs, limitations and abilities. This adaptive ITS model uses real-time behavioral and physiological sensing to allow the ITS to “perceive” affective state and project changes in beliefs, desires and intentions.

**Assessing and modeling affect**

Given the importance we have placed on assessing affect in our adaptive ITS model, the following defines affect and reviews methods to assess and model affect.

**Contrasting personality, mood and emotions**

Personality has long-term affect and reflects individual differences in cognitive and social processes. The Five Factor Model (FFM) of personality (McCrae & John, 1992) defines personality traits in terms of openness, conscientiousness, extraversion, agreeableness and neuroticism; all of which tend to be relatively stable in adult populations. Personality moderates the speed and degree to which mood and emotions emerge and dissipate.

Moods generally have a moderate duration as opposed to emotions (short duration) and personality (long duration). Mood has a subtle influence on cognition which may go unnoticed even by human tutors and are of unknown cause so they may be realized as the cumulative effect of a series of emotional events. According to Morris (1989), mood has “a great influence on human’s cognitive functions” and thereby learning. Mood is represented by Mehrabian’s
(1996) Pleasure-Arousal-Dominance (PAD) model. Mood moderates emotional change and may be influenced by series of emotional events.

Emotions are a psychological state or process that functions in the management of goals and needs of an individual (Broekens & Degroot, 2004). Davidson (1994) argued that “emotions bias action, whereas moods bias cognition.” That infers that emotions may inhibit appropriate actions or cause unintended actions (student behavior), while mood might affect perception and reasoning (input and processing) by the student during the learning process. Emotions are represented by Ortony, Clore & Collins’ (1988) twenty-two emotional states (e.g., joy, pride).

Relationship between mood and personality

Mehrabian (1996) defined mathematical relationships between the PAD dimensions and the FFM factors of personality. We made the assumption that since personality is stable in healthy adults (McCrae & Costa, 1994) and mood is of moderate duration, that mood may be considered to be stable for training sessions of moderate duration (30 minutes to one hour). We tested mood for stability in our experiment and discuss the validity of this assumption in the results section.

Personality assessment methods

Again, since personality traits have been determined to be relatively stable in healthy adults (McCrae & Costa, 1994), we assume that a single evaluation is all that is required to assess the FFM traits. The Big Five Personality Test (John, 2003) is a representative instrument that measures the five FFM dimensions of openness, conscientiousness, extraversion, agreeableness and neuroticism on 0-100 scale.

Mood and emotion assessment methods

Mood and emotion assessments tend to fall into one of three categories (self-report instruments, behavioral indicators and physiological indicators) and may be combined to provide a more accurate assessment of the student’s affective state. Lang’s (1980) Self Assessment Manikin (SAM) is a self-report instrument composed of a nine-point Likert scale for each mood dimension (pleasure, arousal and dominance). Affective prediction methods include the use of: behavioral measures (e.g., facial landmark sensing, keystroke and mouse movement detection); physiological measures (e.g., heart rate inter-beat interval sensing or galvanic skin response); and other student data (e.g. demographics). Examples of related affective computing research discussed below.

Neji and Ben Ammar (2007) investigated an ITS that included a virtual character and facial sensors. In addition to the two-way conversational input and output, the tutor inferred the emotional state of the student through a machine vision system which sensed changes in distances between facial landmarks and classified their expressions as one of six universal emotional states (joy, sadness, anger, fear, disgust and surprise) or a neutral expression. The student’s emotional state was then used in the ITS to determine which tutoring strategy (e.g., sympathizing or non-sympathizing feedback, motivation, explanation, steering). This study took significant steps toward an adaptable tutoring system, but two primary limitations are noteworthy: the cost and portability of the visual system would likely limit the ability of the ITS to support large, distributed populations; and it failed to assess the connection between student affect, the selection of instructional strategies and performance.

D’Mello, Craig, Sullins and Graesser (2006) and D’Mello and Graesser (2007) used frequent conversation patterns to predict affect (i.e. confusion, eureka, frustration) when students emote aloud and their ITS provided feedback, pumps, hints and assertions to influence student’s progress. They noted significant relationships between tutor feedback (negative, neutral or positive) and the student’s affect. Positive feedback by the ITS was a strong positive predictor of eureka, “a feeling used to express triumph on a discovery” (D’Mello et al., 2006). The “directness” of the tutor predicted a negative relationship with confusion. Negative feedback from the tutor was the only significant predictor of frustration. The primary drawback to this approach was the requirement for students to “emote aloud” which has some of same drawbacks as other self-report methods and may be incompatible with students with lower FFM openness scores. Another drawback was the low participant throughput for the experiment based on the labor
intensive nature of the data collection and analysis. Given the variability among students and the associated time to baseline each student, this method is unsuitable as is for training in large-scale, distributed organizations.

Zimmermann, Guttormsen, Danuser and Gomez (2003) used passive methods to capture computer keystrokes and mouse movement as indirect indicators of affect. Their method falls short in that it does not determine any correlation between computer keystrokes or mouse movement and specific affective states (e.g., happy, sad or frustrated). Their approach also fails to follow up in adapting coaching strategies within tutoring systems to the student’s affective state once it is determined.

Yun, Shastri, Pavlidis and Deng (2009) demonstrated passive sensing and interpretation of thermal images to estimate student stress levels. They altered the difficulty levels of game play for users based on singular input from StressCam, which monitors heat dissipation through a thermal imaging-based camera and analysis system. Since stress levels are related with increased blood flow in the forehead and higher blood flow equates to more heat, StressCam passively and continuously senses and interprets thermal images. No specific affective variables were predicted in this research and no evidence was provided regarding the use of this technology for determining instruction. Setup time, system cost and lack of portability may be an issue with this approach.

Whereas many of the machine perception methods reviewed above focus on single data sources (e.g. thermal imaging) to predict affect, others pursued multi-source approaches. McIntyre and Göcke (2007) advocated a multimodal approach to reduce the uncertainty associated with physiological measurements.

Conati and Maclaren (2004) inferred six emotions (joy-distress, pride-shame, and admiration-reproach) of the twenty-two emotions defined in the OCC model (Ortony, Clore & Collins, 1988) by assessing their reaction to meeting/not meeting goals during a computer game. They used a probabilistic model, the dynamic decision network (DDN), which used: personality traits; interaction patterns; electromyography (EMG) which measures muscle activity; and skin conductance to indirectly assess student emotions. The drawbacks to this approach are its accuracy (no significant difference from the baseline accuracy of 50%), and inclusion of only 6 of 22 OCC emotions. It also lacks a student performance assessment to gauge the effectiveness of the selected instructional strategies. Finally, the physiological sensing of valence produced false negatives 55% of the time.

Hernandez, Noguez,Sucar and Arroyo-Figueroa (2006) also developed a Bayesian affective student model based on student personality traits, student knowledge state, student goals and the tutoring situation to infer the same six emotions. A weakness of this approach is in its performance as a tutor as compared to human tutors. Pedagogical actions of their model of instruction only agreed with the pedagogical actions of expert human tutors 57% of the time.

Based on approaches cited in the literature and the adaptive tutoring model discussed above, we decided to pursue a passive approach to detecting mood in the context of a self-paced training environment with an embedded tutor. We chose to assess the influence of student state variables (e.g., competency) and student behaviors as predictors of mood state.

**Hypothesis under test**

Three hypotheses were evaluated in this research. In reviewing the literature, there was little beyond self-report methods used as indicators of student mood. Given the obtrusiveness of self-reporting methods and concerns that frequent assessments would interfere with the learning process, we determined to assess passive predictors of mood and performance, and to test the assumption of mood as a constant during training.

**Hypothesis “A”: Mood prediction**

In hypothesis A, we considered predictors of student mood variables. Mood variables were selected as dependent variables over other affective variables (personality or emotion) based on their duration, influence and cause. Mood was also selected because of its influence on cognition. Davidson (1994) argued that “emotions bias action, whereas moods bias cognition” which indicated that mood would likely affect perception and reasoning during learning.
Based on this connection, we considered the importance of accurately predicting mood variables and posed the following hypothesis: “student state variables (e.g., amount of sleep, energy level, a priori knowledge level and interest in the topic), student action variables (e.g., mouse movement rates and control-selection rates) and student performance are predictors of student mood variables”.

**Hypothesis “B”: Mood stability**

In hypothesis B, we considered the assumption that mood is relatively constant over moderate periods of time (e.g., a training session lasting thirty to sixty minutes). We considered the minimum number of mood assessments required to accurately reflect the student’s mood and still maintain the continuity of the learning process. Would only one or two assessments be required or would mood have to be assessed more frequently or even continuously? To determine stability of mood, we posed the following hypothesis: “student mood variables (pleasure, arousal and dominance) are generally constant during a 30-60 minute training session”.

**Hypothesis “C”: Performance prediction**

In hypothesis C, we considered predictors of student performance and posed the following hypothesis: “student state variables, student action variables and student mood variables are predictors of student performance”.

**Experimental methods**

**Participants**

Cadets from the United States Military Academy (USMA) were the target population for this study. The experimental design focused on the target population with low-moderate competence in tactical combat casualty care (TC3). One hundred-thirty one (131) cadets participated in the study which produced 124 instances of usable data. Of the 124 participants, 108 were males (age M = 18.79, SD = 0.98) and 16 were females (age M = 18.38, SD = 0.62). The 124 participants in this study are approximately 17.7% of the available participants in this population.

**Experimental procedure**

Our experimental approach is based on a case study which involved instructing cadets about the elements of hemorrhage control using a multi-media training package called Tactical Combat Casualty Care (TC3). The content of this study was delivered to the participants via a multimedia presentation on a laptop computer. The study content included a biographical survey, a TC3 training course, a pre-course and post-course knowledge assessment, a pre-training and a post-training mood assessment, and a TC3 performance assessment. Interactive controls, coaching strategy logic, timestamp events and automatic recording of each participant’s interaction and survey data was accomplished through embedded software. The experimental process is shown in Figure 5.

Student state information (e.g., age, gender, energy level, amount of sleep, first aid experience and interest level in TC3) was collected through a biographical survey. The Self-Assessment Manikin (SAM) survey (Lang, 1980) was administered twice to each participant: once prior to the training course and once again after the completion of the TC3 performance assessment. Knowledge assessments were administered prior to the delivery of instruction and after the delivery of instruction.

The training course focused on acquiring knowledge (facts and principles) while the performance assessment focused on applying knowledge and demonstrating skill through decisions and actions. Figure 6 shows a screen capture from the performance assessment which is time sensitive as the student must appropriately treat the wounds within the approximately four minutes for the soldier to bleed out and die.
Results

Hypothesis “A”: Mood prediction

We posed the following hypothesis: “student state variables, student action variables (e.g., behaviors) and student performance are predictors of student mood variables”. We analyzed specific student state and behavior variables with respect to the three mood variables: pleasure, arousal and dominance.

Final pleasure measurements were positively correlated with self-assessed first aid knowledge, Pearson’s $r(124) = .177$, $p = .049$; TC3 interest level, Pearson’s $r(124) = .178$, $p = .048$; initial pleasure, Pearson’s $r(124) = .272$, $p = .002$; initial arousal, Pearson’s $r(124) = .194$, $p = .031$; initial dominance, Pearson’s $r(124) = .266$, $p = .003$; final dominance, Pearson’s $r(124) = .541$, $p < .001$; control selection rate, Pearson’s $r(124) = -.182$, $p = .043$; and performance, Pearson’s $r(124) = .177$, $p = .012$.

Final arousal measurements were positively correlated with initial pleasure, Pearson’s $r(124) = .352$, $p < .001$; and initial arousal, Pearson’s $r(124) = .507$, $p < .001$. 

Figure 6. Screen capture from TC3 performance assessment
Final dominance measurements were positively correlated with TC3 interest level, Pearson’s $r(124) = .233$, $p = .009$; initial arousal, Pearson’s $r(124) = .177$, $p = .049$; initial dominance, Pearson’s $r(124) = .578$, $p < .001$; final pleasure, Pearson’s $r(124) = .541$, $p < .001$; and performance, Pearson’s $r(124) = .393$, $p < .001$.

Examining the significance of the variables and their influence on each mood variable, we developed the models shown in Tables 1 and 2 via linear regression. Initial pleasure, final dominance and control selection rate explained a significant portion of variance in final pleasure scores, $R^2 = .36$, $F(3, 120) = 22.77$, $p < .001$, $d = .57$. Initial dominance, control selection rate, performance and final pleasure explained a significant portion of variance in final dominance scores, $R^2 = .54$, $F(4, 119) = 35.23$, $p < .001$, $f^2 = 1.18$. The regression effect sizes ($f^2$) indicated large effect for both models per Cohen’s (1992) guidelines.

Table 1. Regression Model of Pleasure

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>95.0% Confidence Interval for B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>(Constant)</td>
<td>2.086</td>
<td>.854</td>
<td></td>
<td>2.453</td>
<td>.016</td>
</tr>
<tr>
<td>Initial Pleasure</td>
<td>.230</td>
<td>.099</td>
<td>.173</td>
<td>2.323</td>
<td>.022</td>
</tr>
<tr>
<td>Final Dominance</td>
<td>.519</td>
<td>.073</td>
<td>.523</td>
<td>7.086</td>
<td>.000</td>
</tr>
<tr>
<td>Control Selection Rate</td>
<td>-.090</td>
<td>.037</td>
<td>-.182</td>
<td>-2.472</td>
<td>.015</td>
</tr>
</tbody>
</table>

Table 2. Regression Model of Dominance

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>95.0% Confidence Interval for B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>(Constant)</td>
<td>-2.130</td>
<td>.330</td>
<td></td>
<td>-2.566</td>
<td>.012</td>
</tr>
<tr>
<td>Initial Dominance</td>
<td>.519</td>
<td>.079</td>
<td>.436</td>
<td>6.531</td>
<td>.000</td>
</tr>
<tr>
<td>Control Selection Rate</td>
<td>.068</td>
<td>.032</td>
<td>.136</td>
<td>2.101</td>
<td>.039</td>
</tr>
<tr>
<td>Performance</td>
<td>.021</td>
<td>.099</td>
<td>.155</td>
<td>2.304</td>
<td>.023</td>
</tr>
<tr>
<td>Final Pleasure</td>
<td>.417</td>
<td>.067</td>
<td>.415</td>
<td>8.233</td>
<td>.000</td>
</tr>
</tbody>
</table>

Hypothesis “B”: Mood stability

We posed the hypothesis: “student mood variables (pleasure, arousal and dominance) are generally constant during a 30-60 minute training session”. A non-directional student’s t-Test ($\alpha = .05$) was used to compare the participant’s mood measurements at the beginning and end of the experiment. This comparison revealed a statistically reliable difference between initial ($M = 6.03$, $SD = 1.22$) and final ($M = 5.36$, $SD = 1.63$) pleasure measurements, $t = 1.98$, $p < .01$, $d = .47$; and a statistically reliable difference between initial ($M = 4.05$, $SD = 1.74$) and final ($M = 5.28$, $SD = 1.64$) arousal measurements, $t = 1.98$, $p < .01$, $d = .73$. For dominance, there was no statistically reliable difference between initial and final measurements.

This indicates that our assumption that mood variables remain constant during a 30-60 minute training session are not supported for pleasure and arousal variables, but may be valid for the dominance variable. The effect sizes (Cohen’s $d$) indicated the research environment had medium effect for pleasure and a large effect for arousal per Cohen’s (1992) guidelines.

Hypothesis “C”: Performance

We posed the following hypothesis: “student state variables, student action variables and student mood variables are predictors of student performance”. We analyzed specific student state, behavior and mood variables with respect to student performance. Performance measurements were positively correlated with initial dominance, Pearson’s $r(124) = .289$, $p = .001$; final pleasure, Pearson’s $r(124) = .224$, $p = .012$; final dominance, Pearson’s $r(124) = .393$, $p < .001$; and mouse movement rate, Pearson’s $r(124) = .251$, $p = .005$. 

110
Examining the significance of the variables and their influence on performance, we developed the model shown in Table 3 via linear regression. *Initial dominance, mouse movement rate* and *final knowledge scores* explained a significant portion of variance in performance scores, \( R^2 = .23, F(3, 120) = 11.96, p < .001, f^2 = 0.30. \) The regression effect size \((f^2)\) indicated moderate to large effect per Cohen’s (1992) guidelines.

**Table 3. Regression Model of Performance**

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>95.0% Confidence Interval for B</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Constant)</td>
<td>32.815</td>
<td>8.317</td>
<td></td>
<td>.000</td>
<td>16.347</td>
<td>49.283</td>
<td></td>
</tr>
<tr>
<td>InitialDominance</td>
<td>2.917</td>
<td>.716</td>
<td>.330</td>
<td>.000</td>
<td>1.600</td>
<td>4.333</td>
<td></td>
</tr>
<tr>
<td>MouseMovementRate</td>
<td>.062</td>
<td>.018</td>
<td>.281</td>
<td>.001</td>
<td>.026</td>
<td>.097</td>
<td></td>
</tr>
<tr>
<td>FinalKnowledge</td>
<td>.247</td>
<td>.081</td>
<td>.245</td>
<td>.003</td>
<td>.087</td>
<td>.407</td>
<td></td>
</tr>
</tbody>
</table>

**Discussion, conclusions and future research**

This research examined whether mood could be passively predicted, whether mood could be considered as a constant and the relationship of mood and other student variables were predictors of performance. The objective was to discover relationships that could expand the capability of ITS student models.

Predictive models of pleasure and dominance were developed via regression analysis. Although it was expected that mouse movement and control selection rates would be direct indicators of arousal, there were no reliable predictors of arousal generated during our experimentation.

Mood, which we expected to be stable over the course of a 30-60 minute training session, was significantly different for pleasure and arousal while student’s dominance was relatively unaffected by their experience in the training research environment. Pleasure, arousal and dominance changed by an average of -25%, 58% and 18% respectively. The significant changes in pleasure and arousal may be directly connected to the tactical combat casualty care scenario that the students were exposed to during the experiment.

While we did discover reliable predictors of performance, we were surprised that previous experience, training and topic interest were not among those predictors of performance. Experience and training were used as indicators of competence in our adaptive training model. Given the generally low competence level of the sample population, we should have predicted that these variables would not be reliable predictors of performance.

Additional research is needed to fully validate the full adaptive tutoring model discussed herein, but relevant results in predicting future mood states (pleasure and dominance) were realized along with significant predictors of performance. A regression model of future pleasure and dominance states was developed that relied upon few predictors (3 for pleasure and 4 for dominance). A regression model of performance was developed and includes only 3 predictors to account for 23% of the variance in performance. This finding passes the “common sense” test as demonstrated by: high initial dominance which was an indicator of trainee comfort level in working in a new domain and resulted in higher performance scores; high mouse movement rates which indicated a comfort level in working quickly with the user interface; and high final knowledge scores which indicated that the trainee’s absorbed the lessons and were ready to apply them in the performance assessment test. The findings in this paper also confirmed findings of Davidson (1994) by linking mood to performance outcomes.

The experimental results reviewed in this article provide much needed production rules to support intelligent tutor instructional strategy decision-making based on student mood attributes and other variables of interest whereas other similar work is focused either on cognitive attributes and/or emotions. Mood was chosen based on its duration which is similar in length (30-60 minutes) as many training sessions. In contrast emotional states may only last a few minutes. Production rules based on mood afford the tutoring system the opportunity to assess and update mood state much less often than production rules based on emotional states.

The results discussed herein are significant in that they were drawn from a significant sample size (power > 0.99). The application of the algorithms developed and discussed in this paper will allow for more accurate classification of
student mood and thereby allow computer-based tutors to optimally match instruction (challenge level, flow and feedback) to the student’s affective and cognitive needs. This will improve the learning effectiveness of computer-based tutors so they may successfully substitute for more observant human tutors.

While statistically relevant relationships were discovered in an experimental context, the models discussed herein will ultimately be deployed as part of a learning system that integrates and evaluates multiple learning variables. How these results operate as part of a learning system is yet to be determined by empirical evaluations over the long term, but their application in tutors has already begun. Additional unobtrusive methods are being evaluated.

The authors recommend that research be undertaken to develop a more robust model of the “affective sensing-assessment-instructional strategy selection” process. Additional low-cost, passive sensing techniques should be evaluated on a variety of learning and even mobile learning platforms to push the limits of student modeling and instructional strategy selection. We see another research challenge in the sensing of multiple, simultaneous emotions and we also recommend the continued evaluation of techniques to assess performance, beliefs and competency. We also see the need to enhance predictive models through more continuous measures vice limited discrete measures at the beginning and end of key events. Finally, we see the need for a standard set of instructional strategies that are topic independent so they can be implemented across intelligent tutoring systems.

References


Integrating Algorithm Visualization Video into a First-Year Algorithm and Data Structure Course

Pilu Crescenzi¹, Alessio Malizia², M. Cecilia Verri¹, Paloma Díaz² and Ignacio Aedo²
¹Department of Systems and Computer Science, University of Florence, Italy // ²Computer Science Department, University Carlos III of Madrid, Spain // pileluigi.crescenzi@unifi.it // amalizia@inf.uc3m.es // mariacecilia.verri@unifi.it // pdp@inf.uc3m.es // aedo@ia.uc3m.es

(Submitted February 23, 2010; Revised January 19, 2011; Accepted May 18, 2011)

ABSTRACT
In this paper we describe the results that we have obtained while integrating algorithm visualization (AV) movies (strongly tightened with the other teaching material), within a first-year undergraduate course on algorithms and data structures. Our experimental results seem to support the hypothesis that making these movies available significantly improved students' performances. Moreover, the movies were highly appreciated by the students (both from a comprehensibility point of view and from a usefulness point of view), even though with a low attitude towards the emerging video pod-cast technology. Finally, our results indicate the necessity of integrating the AV movies with audio comment, which seems to be one of the most interesting research questions left open by our study.

Keywords
CS2 course, Algorithm and data structure visualization, experimental evaluation

Introduction
Algorithm visualization (in short, AV) is one of the two main subareas (along with program visualization) of the software visualization research field. In order to distinguish between static and dynamic visualizations of algorithms, in the literature it is often used the term algorithm animation instead of algorithm visualization: in this paper, we will consider the two terms equivalent. By adapting the definition of this latter field given in (Price, Baecker, & Small, 1993), we could say that AV is “the use of the crafts of typography, graphic design, animation, and cinematography with modern human computer interaction and computer graphics technology to facilitate” the human teaching and learning of the design, the behavior, and the analysis of algorithms and data structures. Since its birth (that can be identified with the seminal video of Ron Baecker developed at the beginning of the eighties and successively described in (Baecker, 1998)), AV has been one of the most active research areas within the more general field of computer science education. The 30-minute video, entitled Sorting out sorting, was developed with the assistance of Dave Sherman within the Dynamic Graphics Project of the University of Toronto in 1981, and successively included in SIGGRAPH Video Review 7, 1983. In particular, researchers alternated between the proposal of new tools and languages for developing and distributing algorithm visualizations (such as the ones described in (Rössling & Freisleben, 2002; Naps, Eagan, & Norton, 2000; Shenerson & Tal, 1997; Crescenzi & Nocentini, 2007; Laakso, Myller, & Korhonen, 2009)), and the experimental evaluation of the efficacy of these tools from a pedagogical point of view. This latter line of research culminated in the meta-study described in (Hundhausen, Douglas, & Stasko, 2002), which examined from a more abstract point of view 24 experimental studies performed during the nineties. In a further effort to understand these experimental results and to explore the role of visualization and engagement in computer science education, a framework for experimental studies of AV effectiveness has been proposed in (Naps et al., 2003), based, beyond other things, on an engagement taxonomy including six different forms of student engagement with AV technology.

In this paper, we focus our attention on the full integration of AV movies within the teaching material distributed to the students of a first-year algorithm and data structure course. Our goal is to experimentally validate (we refer to experimental validation in general sense and not to controlled experiment) the hypothesis that making AV movies available can significantly improve students' performances, whenever the visualization is strongly tightened with the front lecture and with the lecture notes. To this aim, we have developed eight sorting AVs, which have been distributed (in MPEG-4 format) along with the lecture notes after the front lecture. We have successively analyzed the students' grades obtained while taking a written exam on topics related to the AV movies. Finally, we have asked the students to fill surveys concerning their opinion about the usefulness of the AV movies.
The rationale behind the video format choice is that we also wanted to somehow verify the students' attitude towards the emerging video podcast technology. Indeed, in the last few years several papers (Campbell, 2005; Brittain, Glowacki, Van Ittersum, & Johnson, 2006) have depicted this technology in a quite enthusiastic way, by imagining a world in which students would listen “to a podcast on the drive to school, then reinforcing the day’s learning by listening to another podcast, or perhaps the same podcast, on the drive back home” (Campbell, 2005). The second goal of this paper is, then, to quantitatively and qualitatively verify the attitude of our students towards this near future scenario.

The rest of the paper is organized as follows. In Section II we briefly describe the AV tool we have used in order to produce the AV movies. Sections III and IV report the evaluation experiment carried out in order to evaluate the efficacy of AV movies from a pedagogical point of view. Finally, in Section V we summarize the related work and in Section VI we provide some conclusions and future work.

**AIViE**

In order to develop the AV movies, we used an algorithm visualization system, called AIViE, which is a *post-mortem* tool (according to the post-mortem approach, visualization takes place after the algorithm execution and requires some sort of trace files which is recorded during the execution itself) based on the *interesting event* paradigm (see, for example, (Price et al., 1993)). In particular, AIViE includes (1) a visualization player that allows the user to navigate through a visualization, (2) a graphical input developer that allows the user to create new input and, hence, to change a visualization, and (3) a Java class library, which allows a programmer to create new visualizations on the ground of the following data structures: arrays, matrices, lists, binary trees, graphs, queues, and stacks.

The AIViE distribution we used (Crescenzi, Gambosi, Grossi, Nocentini, & Verdese, 2007) included 65 visualizations of algorithms and data structures. This collection of visualizations is physically included in the distribution and can, hence, be used off-line, that is, without the necessity of being connected to a world wide web server, as in the case of JHAVE, another AV environment which comes with several algorithm and data structure visualizations (Naps, 2005).

The user interface of AIViE can be fully localized: indeed, every graphical and textual component of the interface is associated with several properties whose value is specified within a language-dependent text file. Currently, the interface of AIViE is available in Italian and in English: creating a new localization is a matter of a few time translating task.

Almost all the visualizations included in the AIViE distribution provide the possibility of showing the pseudo-code of the algorithm, which is written in a pseudo-C language. Moreover, by specifying the value of several properties contained into language-dependent text files, all the visualizations can be easily customized from a graphical point of view, and they all support localization: currently, all the visualizations are available in Italian and in English.

The user can create new visualizations of any of the algorithms included in the AIViE distribution by creating new inputs by means of the graphical input developer, which is part of the AIViE framework. Moreover, the Java class library allows a programmer to quite easily construct new visualizations of new algorithms: in the last three years, this feature has been extensively used and more than forty new visualizations have been produced by almost eighty second year undergraduate students. Even in this case, as far as we know, this is one of the largest collection of visualizations constructed by students by means of the same AV tool.

AIViE is currently supported by an Italian text book on algorithms and data structures (Crescenzi, Gambosi, & Grossi, 2006), which is, as far as we know, the first book which fully integrates AV technology into the theoretical design and analysis of algorithms and data structures. Indeed, very few examples are printed within the book, while most of the time the explanation of an algorithm refers to the visualization of the algorithm itself. The text book has been adopted in the universities of Florence, Pisa, Siena, and Verona in Italy and it is suggested in almost twenty other Italian computer science undergraduate programs.
Quantitative evaluation

In this section we present the results of the experimental study we have performed in order to determine whether the distribution of AV movies along with the lecture notes improves students' performances. In particular, we intended to demonstrate that viewing AV movies can improve students' grades by making use of a scenario-based design approach (Rosson & Carroll, 2002). To this aim, we have designed an experimental scenario among the students of a first-year algorithm and data structure course, which was made of two intermediate exams for the course. We have then performed a statistical analysis of the grades obtained by the students while taking these intermediate exams. At the same time, we asked the students to fill surveys in order to express their opinion about the usefulness of the movies and the technological aspects of the visualizations.

The course

Algoritmi e Strutture Dati is a first-year course of the Computer Science undergraduate program at the Science Faculty of the University of Florence. The topics covered by the course are the following: algorithm complexity analysis, elementary data structures (arrays and lists), abstract data structures (stacks, queues, priority queues, trees), search algorithms (binary search, binary search trees, AVL and 2-3 trees, tries, random search), sorting algorithms (elementary algorithms, quicksort, mergesort, heap sort), graphs (representation, search algorithms, MST). The course corresponds to 12 ECTS credits (the European Credit Transfer and Accumulation System is a standard for comparing the study attainment and performance of students of higher education across the European Union, with 12 credits being approximately 100 front lecture hours). The individual student's grade is decided on the ground of four distinct exams: a first intermediate written exam is taken after the first part of the course (that is, up to the search algorithms), a second intermediate written exam is taken at the end of the sorting algorithm part, while an oral exam and a laboratory exam (to be arranged along with the Programmazione course) are taken at the end of the course. Programmazione is a programming course which also corresponds to 12 ECTS credits: even though the laboratory exam is arranged along with Algoritmi e Strutture Dati, the grades assigned to the students in the two courses are independent. The course (as almost any course of the undergraduate program) uses the Moodle platform (Cole & Foster, 2007) in order to distribute the learning material and to manage all the related web activities.

The students

Almost 200 students signed to the course on the Moodle platform. However, only 91 students took the first intermediate exam: 54 of them also took the second intermediate exam. Our sample was formed by these latter 54 students. The 54 students represent one fourth of the registered students but this is a quite normal trend in Italy for this kind of first-year courses, in which there is a high abandoning rate.

The AV movies

Eight movies were produced and distributed, related to the three more advanced sorting algorithms (that is, quicksort, mergesort, and heap sort). The movies were developed by making use of AlViE and of the ScreenFlow software (Telestream, Inc, 2009) (in order to capture the screen activity and to produce MPEG-4 movies). They were distributed along with the lecture notes just after the front lectures. All movies include the algorithm pseudo-code (taken from the lecture notes): moreover, as a consequence of the comment of one student, the last six movies also visualize the values of most of the auxiliary variables used by the pseudo-code. The movies can now be downloaded starting from the web page of AlViE (Crescenzi et al., 2007).

Independent variable

There was only one independent Boolean variable, that is, whether the students viewed or not some of the movies that were distributed. By using the Moodle platform, we could trace the movies downloads: however, downloading a movie does not imply viewing it. Hence, since the students were not forced to download and to see the movies, in order to quantify this variable, we asked the students to fill a survey for each of four groups of videos, concerning the
usefulness of the movies and their technological aspects (see Appendix A). The first group of two videos concerned the basic quicksort algorithm, the second group of two videos was related to two modified version of the quicksort algorithm, the third group contained only one video of the mergesort algorithm, and the last group of three videos concerned the heap sort algorithm.

Note that, since the experimental set-up is scenario-based and not a controlled experiment, we could only infer our conclusions about the independent variable values assuming it equal to true if and only if the student filled a sufficient number of surveys to guarantee that he/she viewed at least two movies. Note that a couple of students filled a survey without downloading any movie: this is not a contradiction, since it is reasonable to assume that movies were also passed among the students.

According to the above criterion, the 54 students have been divided into two groups: the first group contained the 26 students that watched at least two movies, while the second group contained the remaining 28 students that did not watch any movie at all (or, at least, did not fill any survey): interestingly, the sample is almost split into two halves. In the following, we will denote with \( Y \) and \( N \) the first and the second group, respectively.

**Dependent variable**

Even in this case, there was only one dependent variable, that is, the grade obtained by the student while taking the second intermediate exam. In particular, this exam contained the following five questions (the students were given up to two hours to complete the exam):

- **Q1.** What is a stable sorting algorithm? Give at least three examples of stable sorting algorithms.
- **Q2.** Find a permutation of the first 10 integer positive numbers on which the quicksort algorithm, with the pivot chosen as the median of three, executes the maximum number of comparison.
- **Q3.** Sort the array 15, 5, 12, 10, 20, 23, 18, 13, 3, 6, 7, and 19 by using the insertion sort algorithm.
- **Q4.** Sort the array of the third question by using the quicksort algorithm with the pivot chosen as the rightmost element.
- **Q5.** Sort the array of the third question by using the heap sort algorithm.

Each question is evaluated by means of a grade between 0 and 6. Hence, the dependent variable is a number between 0 and 30.

**Data analysis**

We denote by \( \mu_Y \) (respectively, \( \mu_N \)) the mean of the grades of the students in \( Y \) (respectively, \( N \)). Moreover, we denote by \( \sigma^2_Y \) (respectively, \( \sigma^2_N \)) the variance of the grades of the students in \( Y \) (respectively, \( N \)). Finally, we denote by \( \mu_Y^{1/2} \) (respectively, \( \mu_N^{1/2} \)) the median of the grades of the students in \( Y \) (respectively, \( N \)). The values of these measures are summarized in the following Table 1 (30 was the maximum grade).

<table>
<thead>
<tr>
<th>Measure</th>
<th>( \mu_Y )</th>
<th>( \mu_N )</th>
<th>( \sigma^2_Y )</th>
<th>( \sigma^2_N )</th>
<th>( \mu_Y^{1/2} )</th>
<th>( \mu_N^{1/2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>23.9</td>
<td>21.1</td>
<td>4.2</td>
<td>6.3</td>
<td>24.0</td>
<td>20.5</td>
</tr>
</tbody>
</table>

We could not assume that \( \sigma^2_Y \) and \( \sigma^2_N \) are equal, since the results of an F-test to compare the standard deviations suggests that there may be a significant difference between them. For this reason, we did not run a Student’s t-test analysis to compare the means: instead, we ran a Mann-Whitney-Wilcoxon test (Dix, Finlay, Abowd, & Beale, 2003) to compare the medians. This test, which is one of the best-known non-parametric significance tests, is constructed by combining the two samples, sorting the data from smallest to largest, and comparing the average ranks of the two samples in the combined data. By applying the test to our data, we obtain a \( P \)-value equal to 0.043: since this value is less than 0.05, we can conclude that there is a statistically significant difference between the medians at the 95% confidence level. This implies that the difference of the means is significant (see also Figure 1(a)). Even if the statistical results are not strictly related to a controlled experiment, by considering the settings proposed at the beginning of this section we can conclude that the average grade was significantly higher in the case of students that
watched the video material and, in particular, in the case of the last 3 questions where students scored better results distributed uniformly.

![Figure 1](image)

*Figure 1.* (a) Students that did not watch any video (No_video) against students that did watch at least two videos (Video): *X*-axis presents grades, while *Y*-axis has two values if videos have been watched or not. (b) Students using videos have more uniformed grades that tend to be higher than students that did not watch any video (whose grades are more scattered and seem to depend only on students’ skills).

A flattening effect that uniforms results of students that watched videos clearly appears in Figure 1(b) (Exercise 3, 4, and 5), where the *X*-axis presents grades, and the *Y*-axis has two values if videos concerning questions 3, 4, and 5 have been watched or not. Thus students using videos have more uniformed grades that tend to be higher than students that did not watch any video (whose grades are more scattered and seem to depend only on students’ skills).

**Further results**

In order to be sure that the above result is not due to the fact that students in *Y* are consistently better than students in *N*, we performed the same data analysis by using as dependent variable the grade obtained by the students while taking the first intermediate exam, which consisted of questions not related to the topics explained in the videos. In this scenario such evaluation can be comparable to a control group in a controlled experiment even if in our case conclusions cannot be as strong because of the open scenario (i.e. no time on task has been measured). Referring to the previously introduced notation, the measure values are summarized in Table 2.

<table>
<thead>
<tr>
<th>Measure Value</th>
<th>µ₉₀</th>
<th>µ₁₅</th>
<th>σ₀₀</th>
<th>σ₁₅</th>
<th>μ₁/₂</th>
<th>μ₁/₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>23.7</td>
<td>22.0</td>
<td>4.3</td>
<td>5.1</td>
<td>23.0</td>
<td>23.0</td>
</tr>
</tbody>
</table>

By running the Mann-Whitney-Wilcoxon test on these data, we obtain a *P*-value equal to 0.31: since this value is greater than 0.05, we can conclude that there is no statistically significant difference between the medians at the 95% confidence level. This means that the difference of the means is not statistically significant. In other words, differently from what we could do in the case of the second intermediate exam, in this case we cannot reject the null hypothesis and hence we cannot draw conclusions about students’ level of preparation.

As a further support to this statement, we have also compared the grades obtained by the students, while taking the second intermediate exam, relative to the first two questions (which were more theoretical) with the ones relative to the last three questions (which were more practical). Once again, referring to the same notation, the measure values are summarized in Table 3.

By running the Mann-Whitney-Wilcoxon test on the first two question data, we obtain a *P*-value equal to 0.369: since this value is greater than 0.05, we are not allowed to conclude that there is a statistically significant difference between the medians at the 95% confidence level. On the other hand, by running the Mann-Whitney-Wilcoxon test on the second three question data, we obtain a *P*-value equal to 0.002: since this value is less than 0.05, we can conclude that there is a statistically significant difference between the medians at the 95% confidence level. In other words, as we could expect, the students performance improvement by watching videos is mainly related to more practical questions, in which the simulation of an algorithm was required. Moreover if we observe the variance in the
case of the first two questions and of the other three, then we can notice an interesting phenomenon: the variance decreases ($\sigma^2_Y = 3.7$ in the case of the first two questions and $\sigma^2_Y = 1.0$ in the case of the last three). This could be explained by the fact that, as already noticed, the use of videos has a sort of flattening effect on the learning curve.

Table 3. Students’ grades for the first intermediate exam split in first part (two questions) and second part (three questions): (Y) watched at least 2 videos, (N) did not watch any video.

<table>
<thead>
<tr>
<th>Measure</th>
<th>$\mu_Y$</th>
<th>$\mu_N$</th>
<th>$\sigma^2_Y$</th>
<th>$\sigma^2_N$</th>
<th>$\mu^Y_{1/2}$</th>
<th>$\mu^N_{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>First two questions</td>
<td>6.2</td>
<td>5.4</td>
<td>3.7</td>
<td>4.1</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Last three questions</td>
<td>17.7</td>
<td>15.6</td>
<td>1.0</td>
<td>3.4</td>
<td>18.0</td>
<td>17.5</td>
</tr>
</tbody>
</table>

Qualitative evaluation

As we have already said before, the students were asked (but not forced) to express their opinion about the AV movies by filling one or more surveys. Just before the second intermediate exam, we collected the data contained in the 55 surveys filled by 31 students (note that some students clearly filled more than one survey). The results are summarized as follows:

- Only 5 students (that is, 16%) viewed at least one movie on an iPod/iPhone or on another PDA device.
- Of the above 5 students, one judged the video resolution as very good, three judged it as good, and one judged it as bad.
- In 12 surveys (that is, 22%) the visualization was judged very comprehensible. In all the other surveys, it was judged comprehensible.
- In 21 surveys (that is, 38%) the visualization was judged very useful, in 33 it was judged useful, and only in 1 it was judged unuseful.
- In 40 surveys (that is, 73%) the integration of an audio comment was judged useful.
- In 19 surveys over 32 (that is, 59%) the visualization of the auxiliary variable was judged very useful, in 10 surveys (that is, 31%) it was judged somehow useful, in 2 surveys (that is, 6%) it was judged maybe confusing, and in the remaining survey (that is, 3%) it was judged completely unuseful and confusing.

From these results, it is clear that the students appreciated the AV movies, both from a comprehensibility point of view and from a usefulness point of view. Interestingly, it seems that the video pod-cast technology is not very spread, and that the students prefer to see the movies on their desktop/laptop displays. Indeed, it could also be that most of the students do not even own an iPod/iPhone or another PDA device.

Related work

As we already said in the introduction, several experimental studies have been carried out concerning the pedagogical efficacy of the AV technology (most of these studies are surveyed in (Hundhausen et al., 2002)). As far as we know, however, our work is the first attempt to evaluate the strong integration of this technology within an algorithm and data structure course in a quite transparent way: as we already specified, the students were not forced to see the movies neither to fill the surveys. We also emphasize the fact that our integration of the movies was very tightened with the course contents: indeed, the movies show exactly the same pseudo-code that was explained by the teacher during the front lectures and that was included in the lecture notes. As far as we know, the only similar example of such a systematic integration within an undergraduate official course is described in (Crescenzi & Nocentini, 2007).

For what concerns the use of video pod-casts within a learning environment, apart from the two papers already cited in the introduction, it is worth mentioning (Lazzari, 2009): in this paper the author describes an academic experience of pod-casting, which involved a group of students of a course on multimedia communication and human-computer interaction. Finally, in (Hürst, Lauer, & Nold, 2007) the efficacy of algorithm visualization on mobile devices is
studied by setting up an empirical experiment with 16 advanced students of computer science. In this experiment the independent variables were the kind of used device and the presence of an audio comment.

Conclusions

In this paper we have described the results obtained while integrating AV movies within a first-year undergraduate course on algorithms and data structures. These results, which are promising both from a student performance point of view and from a student appreciation point of view, can be summarized as follows:

- AV movies seem to be a proficuous way to get the students to invest their time while learning an algorithm. Indeed, after one month of front lectures, students who viewed and reviewed at least two movies performed in a statistically significant way (with respect to the Mann-Whitney-Wilcoxon test) better than the other students, while taking an intermediate exam fully devoted to the visualized sorting algorithms. Clearly, the same effect could have happened if one group of students would have read a text book, while the other group did nothing. However, we believe that AV movies have some motivating factor that we cannot forget: our results support the hypothesis that AV movies may be an effective way of helping students use time well (which is one of the seven principles of good practice in higher education indicated in (Chickering & Gamson, 1987)).

- The above performance difference is mainly concentrated on simulation exercises. Students who viewed the movies performed better on these exercises while answering on more theoretical questions did not result in any statistical significant difference. Observe that this result somehow supports the hypothesis that making AV movies available is the main justification for the better performances, since they get the students to invest more time in understanding the behavior of an algorithm.

- Students preferred to see the movies on their desktop/laptop display rather than on iPod or another PDA. Indeed, only a negligible percentage of students viewed the movies on these latter devices. It is worth noting, however, that we did not ask the students to specify whether they owned such a device: hence, this result could be simply due to the fact that these devices were not very popular among the analyzed cohort of students.

- Basically all students considered the AV movies a useful tool for reviewing and studying the material covered by the lecture notes. As a consequence of this result, the integration of AVs into a learning management system such as Moodle seems to be a very interesting line of future research. Indeed, this possibility has already been explored in (Rößling & Vellaramkalayil, 2009), where the authors present a Moodle activity that is able to incorporate animations created in the ANIMAL system (Rössling & Freisleben, 2002).

- Many students stated that the integration of an audio comment into the movies would have been useful. This result seems to be consistent with the experiment performed in (Hürst et al., 2007)) and suggests an interesting research direction. The main problem, in this case, is twofold. From a technical point of view, it is not easy to automatically add an audio comment to a movie, if we want to maintain the internationalization of our software. From a visualization efficacy point of view, instead, an automatically produced audio comment could turn out to be boring and, hence, not very effective. For this reason, we think that satisfying this students' request might turn out to be quite challenging.

As a consequence of the results described in this paper, the new version of AlViE is now capable of automatically recording visualizations and of saving them in different video formats: we hope that this feature will turn out to be a good tool for study strategy (as also suggested in (Yi-Chuan & Cifuentes, 2006)). Besides the experimental results we have described in this paper, there are at least two other main reasons that justify this extension. First of all, movies can help algorithm visualization producers while advertising and distributing new algorithm visualizations. By using movies, there is no need to download a specific software: the final user can see the visualization (even on a browser) and subsequently decide whether it is worth using it and changing it (for instance, by executing the algorithm with different inputs). In this latter case, of course, the necessary software has to be downloaded. Secondly, producing movies can turn out to be a motivating way for students to present their visualizations. On the ground of the first author's experience, it seems that students like to construct and to present their own algorithm visualizations: we think that if they can also distribute them by means, for example, of a video pod-cast directory, then they might even be more motivated in realizing an effective visualization (which should turn out in a better understanding of the algorithm).

From a more pedagogical point of view, finally, an interesting line of research would be to explore the effects of offering the possibility of interacting with the AVs (for instance, by producing new visualizations with different inputs). In particular, during the next academic year we plan to perform experiments similar to the ones described in
this paper in order to evaluate the efficacy of fully integrating this possibility. Once again, our hypothesis will be that whenever the algorithm animations are strictly tightened to the other course material then the possibility of creating new visualizations with different inputs will further improve student performances.

Acknowledgments

The first author is grateful to Cliff Shaffer for several useful e-mail discussions. We thank the anonymous referees for many valuable comments, which improved the presentation of our results.

References


## Appendix A

### The survey questions

In the Table A.1 we show the questions included in the four surveys (the last four questions had to be answered only if the student answered yes to the second question).

Table A.1. Questionnaire presented to the students.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you downloaded the movies?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Have you seen the movies?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>If you answered yes to the previous question, which device have you used?</td>
<td>Desktop or laptop screen</td>
</tr>
<tr>
<td></td>
<td>iPod or iPhone</td>
</tr>
<tr>
<td></td>
<td>Other PDA</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td>If you used iPod/iPhone or other PDA, how was the resolution?</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>Bad</td>
</tr>
<tr>
<td></td>
<td>Very bad</td>
</tr>
<tr>
<td>How was the visualization from a comprehensibility point of view?</td>
<td>Very comprehensible</td>
</tr>
<tr>
<td></td>
<td>Comprehensible</td>
</tr>
<tr>
<td></td>
<td>Not comprehensible</td>
</tr>
<tr>
<td></td>
<td>Completely incomprehensible</td>
</tr>
<tr>
<td>How was the visualization from a usefulness point of view?</td>
<td>Very useful</td>
</tr>
<tr>
<td></td>
<td>Useful</td>
</tr>
<tr>
<td></td>
<td>Not useful</td>
</tr>
<tr>
<td></td>
<td>Completely useless</td>
</tr>
<tr>
<td>How was the visualization of the auxiliary variables? (Not in the first survey)</td>
<td>Very useful</td>
</tr>
<tr>
<td></td>
<td>Somehow useful</td>
</tr>
<tr>
<td></td>
<td>Might be confusing</td>
</tr>
<tr>
<td></td>
<td>Completely useless and confusing</td>
</tr>
<tr>
<td>Would the integration of an audio comment have been useful?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>
An Assessment of E-training Effectiveness in Multinational Companies in Malaysia

Thurasamy Ramayah, Noor Hazlina Ahmad* and Tan Say Hong
Universiti Sains Malaysia, 11800 Penang, Malaysia // ramayah@usm.my // hazlina@usm.my // sayhongtan@yahoo.com

*Corresponding author

(Submitted March 12, 2010; Revised October 11, 2010; Accepted April 18, 2011)

ABSTRACT
E-training has developed into a revolutionary way of learning in Malaysian organizations due to rapid growth in information technology infrastructure. The present study endeavors to determine the critical factors that influence e-training effectiveness in multinational companies (MNCs) in Malaysia. By integrating Technology Acceptance Model (TAM), Delone and McLean information system success model and Expectation Confirmation Model (ECM), the study tested the causal pathway of the factors that could predict the effectiveness of e-training. Data from 163 employees in multinational companies (MNCs) with prior e-training experience was obtained via survey method. The result revealed that motivation to learn, management support and organization support were significant predictors of user satisfaction, however, self-efficacy, ease of use, and content of training were not significant. User satisfaction was also found not to influence net benefit but was a strong predictor of intention to continue usage as hypothesized. The value of this study lies in its effort to propose to practitioners that e-training system requires substantial investment not only at the development and implementation phase, but also the subsequent maintenance.

Keywords
E-training effectiveness, Technology Acceptance Model (TAM), DeLone and McLean Information System success model, Expectation-Confirmation Model (ECM)

Introduction
The advancement of information technologies has contributed to the exponential growth in corporate e-learning (or e-training) in recent years. This revolution enables employees to obtain an intimate learning experience without attending a brick-and-mortar facility. As organizations strive to enhance their competitiveness by constantly promoting continuous learning culture, online training continues to grow in popularity as organizations strive to better meet their immediate and strategic needs for a flexible, well-trained workforce (Kosarzycki et al., 2002).

The development of e-training is indeed not a new phenomenon in Malaysia. Firms in Malaysia have been hopping on to the e-learning bandwagon, hoping to reap benefits from this upcoming trend. However, there are issues and challenges with the implementation of e-learning that need to be avoided or resolved in order to enhance the user satisfaction and e-training effectiveness. Ali (2004) has identified seven issues and challenges in implementation of e-learning in Malaysia. These include lack of awareness amongst population, low adoption rate, bandwidth and connectivity limitations, computer illiteracy, lack of quality e-learning content, difficulty in engaging learners online and language barrier. Limitation in bandwidth and connectivity will affect the response time of the e-learning system among Malaysian organizations (Ali, 2004). Slow response from the e-learning system will create frustration and boredom among users and leads to low satisfaction (Almutairi & Subramanian, 2005).

Hence, careful assessment and evaluation of real organizational learning benefits derived from e-training are required to justify the investment. Some of the important questions to raise are; “Are the employees satisfied with the e-training system provided by the organizations?”, and more importantly, “Are they willing to continue pursuing e-training in the future and are the benefits articulated realized?” Understanding of the critical factors that lead to user satisfaction and effective training outcomes are fundamental for organizations to reap the benefits of e-training. The present study therefore endeavors to determine the critical factors that influence e-training effectiveness in organizations.
E-training in organizations

E-training resembles e-learning in many ways especially in terms of the methods of delivery and technology used, except that it refers to much shorter time frame of learning which usually is specifically designed to achieve a certain learning objective or skill. Typical types of e-training are video conferencing and web-based training. These technologies purport to deliver learning which is “new, better, cheaper, and faster” (Bardach, 1997; Taylor, 2002) than traditional classroom methods. For the purpose of this study, e-training refers to any type of training provided in organizations via electronic media which include self-paced learning from Intranet, learning from CD-ROM at work, training provided by instructors live through Webcast, recorded sessions of past webcast trainings available to employees and others.

The benefits of e-learning have been widely discussed including cost-effective, timely content, and access flexibility (Hong, Lai, & Holton, 2003; Lorenzetti, 2005; Rosenberg, 2001). Bouhnik and Marcus (2006) have specifically stated that e-learning has four advantages: (1) freedom to decide when each online lesson will be learned, (2) lack of dependence on the time constraint, (3) freedom to express thoughts and (4) accessibility to the course’s online materials.

Despite all the perceived benefits of e-learning, research indicates that a high rate of students who commence e-learning courses do not finish them (Dutton & Perry, 2002). Many are dissatisfied with the e-learning experience. A study by Loh (2007) on e-learning effectiveness in the manufacturing industry in Malaysia found that system quality, information quality, and perceived usefulness influenced e-learning effectiveness positively. Another study by Md. Hashim Nordin (2007) that investigates the factors contributing to continuance intention and user satisfaction in e-training has found that information quality and system quality were important factors leading to increase in usage and user satisfaction of e-learning in workplace. Higher level of satisfaction was also found to increase the intention to continue use.

Underpinning theories

Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM), developed by Davis et al. (1989) intends to measure, predict, and explain user acceptance of information technology (refer to Figure 1). TAM theorizes that perceived usefulness and perceived ease of use determine users’ behavioral intention and actual usage. The causal relationships among these constructs have been validated empirically in many studies. Over the last two decades the TAM has been one of the most influential research models in studying the determinants of IT usage (Chau, 2001).

Delone and McLean Information Systems Success (ISS) Model

DeLone and McLean’s Information System Success Model (1992) touches on six dimensions of system success which include system quality, information quality, use, user satisfaction, individual impact, and organizational impact. System quality and information quality singularly and jointly affect both use and user satisfaction. Intention to use (use) has positive and negative effects on the degree of user satisfaction and vice versa. Both use and user satisfaction influence an individual, which will eventually impact on an organization. An updated model of ISS
DeLone and McLean (2003) added “service quality” measure as a new dimension of the IS success model and grouped both the individual and organization impact measures into a single impact or benefit category called “net benefit”.

**Expectation-Confirmation Model**

The Expectation-Confirmation Theory (ECT) asserts that consumers’ intention to repurchase a product or service is significantly influenced by their prior experience with that product or service (Anderson & Sullivan 1993; Oliver 1980). Lower expectation and/or higher perceived performance may lead to a greater confirmation, which results in positive influences to customer satisfaction and continuance intention. Reversing the relationship would cause disconfirmation, dissatisfaction, and discontinuance intention. Confirmation refers to the expectation-performance discrepancy (Hayashi, Chen, Ryan & Wu, 2004). Bhattacherjee (2001) suggests that the IS users’ continuance decision is similar to consumers’ repurchase decision because both decisions (1) follow an initial (acceptance or purchase) decision, (2) are influenced by the initial use (of IS or product) experience, and (3) can potentially lead to an ex-post reversal of the initial decision.

**Trainee dimension: Computer self-efficacy and motivation to learn**

According to Compeau and Higgins (1995), computer self-efficacy is an individual’s belief in his or her ability to perform specific computer tasks. It is believed that individuals having high level of computer self-efficacy are more likely to engage in computer tasks and demonstrate persistence in completing computer tasks despite difficulties (Johnson, Lester & Ferguson, 2001). Likewise, Lim et al. (2007) found that the higher the trainee’s computer self-
efficacy regarding online training, the higher their learning performance. Evidently, computer self-efficacy is important in the e-training environment. Employees who have low self-efficacy often find themselves having problems using the e-training system, as they get minimal guidance from trainers compared to a face-to-face training environment. Therefore, employees with higher computer self-efficacy are believed to have higher satisfaction on the e-training system. Based on this contention, it is postulated that:

H1: High computer self-efficacy is positively related to user satisfaction.

Apart from computer self-efficacy, learner’s motivation is deemed important in e-training. Motivation is defined as “the specific desire of the trainee to learn the content of the training course” that is characterized by both the trainees’ enthusiasm for learning and the trainees’ persistence in attempting to learn the materials when the content becomes more difficult (Noe, 1986, p. 743). In a study by Mathieu, Tannenbaum, and Salas (1992), trainees showed more positive emotional responses when they had higher motivation. Similarly, Mathieu and Martineau (1997) demonstrated that trainees who are motivated tend do well and are more satisfied with the training. Thus, it is conjectured that:

H2: High motivation to learn is positively related to user satisfaction.

**Content dimension: Ease of use and contents of training**

Perceived ease of use is defined as “the degree to which a person believes that using a particular system would be free of physical and mental effort” (Davis, 1989, p. 320). Roca and Gagné (2008) found that perceived ease of use was a significant predictor of e-learning continuance intention. According to Lim et al. (2007), the main factor influencing trainees’ learning performance depends on how easy it is to use or to access a site. Easy navigation through the system and clear online help menu as well as instruction help employees to get the information they need easily and quickly. This will enhance their satisfaction in using the e-training system, which will in turn, increase the continuance intention and the net benefits to individual.

DeLone and McLean (1992), in their information systems (IS) success model suggest that system quality implies an information process system quality based on production of produced information that reflects the accuracy and efficiency of the system. System quality is the measure of the information system process and is said to affect user satisfaction (DeLone & McLean, 1992). It is acknowledged that system quality in information systems success model is substituted for perceived ease of use (Rai et al., 2002; Seddon & Kiew, 1997). A study by Van Slyke, Belanger, and Comunale (2004) that explore the factors that may impact consumers’ decision to engage in Web-based shopping found that perceived complexity (ease of use) has a strong impact on intention to use. Based on the discussion, two hypotheses were formulated:

H3: Ease of use is positively related to continuance intention.
H4: Ease of use is positively related to user satisfaction.

Apart from ease of use in an e-training environment, the contents of training are crucial in the success of any training program. The content has to be up-to-date, relevant, and accurate. Alliger, Tannenbaum, Bennett, Traver, and Shotland (1997) argue that when trainees recognized that the contents were practical, they applied knowledge and skills from the training to their real work. According to Negash, Ryan and Igbaria (2003), information quality is defined as a function of the value of the output produced by a system as perceived by the user. Measures associated with information quality include content variety, complete information, detailed information, accurate information, timely information, reliable information, and appropriate format (Ahn, Ryu & Han, 2004). The information quality dimension evaluates the e-training content whether it has the current, comprehensive, accurate information. In this study, the contents of training refers to the information quality dimension as presented in the DeLone and McLean (1992) information systems (IS) success model. Therefore, it is hypothesized that:

H5: Contents of training is positively related to user satisfaction.
Organization dimension: management and organization support

According to Igbaria, Guimaraes, and Davis (1995), external factors such as organizational characteristics have significant effects on individuals’ intentions to use technology system based on the presence or absence of the necessary skills, opportunities, and resources to use the system. Management support is perceived as one of the important dimensions in an organization training climate which includes top management encouragement, allocation of resources, and instructional development assistance. Kozlowski and Hults (1987) conducted a study that examined the relevance and influence of a training-specific climate dimension. They found strong relationships between seven dimensions of “technical updating climate” (i.e., supervisory support, innovation policies, and job assignments) and individual performance, organizational commitment, and growth satisfaction. Evidently, when the managers show constant emphasis on the importance of continuous learning and tie certain reward to the initiative taken by their employees who actively participate in e-training, employees will feel satisfied and thus, leads to higher training effectiveness. Therefore, it is contended that:

H6: Management support is positively related to user satisfaction.

In addition, previous research has demonstrated that an organization’s training climate is instrumental in preparing individuals for formal development activities and achieving desired learning objectives (Tracey et al., 2001). Lim et al. (2007) found that in an e-training environment, maintaining a consistent learning environment is not significant with transfer performance, but rather with learning performance. When the organization environment emphasizes innovation, and the organization rewards trainees’ self-development efforts, trainees learning performance will be higher. When the organization align their performance management procedures and incentive programs with their employees initiatives in using the e-training system effectively and applying the skills acquired in work, employees will have higher satisfaction in using the e-training system. Therefore, the relationship between organizational support and user satisfaction is equally important in the e-training environment. In light of this, it is hypothesized that:

H7: Organization support is positively related to user satisfaction.

User satisfaction, continuance intention, and net benefits

Previous research has found that individuals who are less satisfied with their experiences are less likely to enroll in future e-learning courses (Lim, 2001; Carswell & Venkatesh, 2002). Similarly, Wang (2003) found a positive relationship between e-learner satisfaction score and the reuse intention of the e-learning systems. Expectation Confirmation theory states that IS continuance intention is positively correlated with satisfaction. A number of research studies on IS satisfaction indicate that IS satisfaction directly effects intention to use and actual use (Davis et al. 1989; Karahanna et al. 1999; Taylor & Todd, 1995). Hayashi et al. (2004) found that the higher the satisfaction level with initial use end users have, the higher their continuance intention in using the virtual learning environments. Liaw (2008) conducted a study to understand how to improve e-learning satisfaction, behavioral intention, and to enhance learning effectiveness on 424 university students in Taiwan. The results of the study showed that perceived satisfaction contribute to the learner’s behavioral intention to use the Blackboard e-learning system. For this study, the relationship between user satisfaction and continuance intention can be tested via the following hypothesis:

H8: User satisfaction is positively related to continuance intention.

Gatian (1994) found that there is a powerful relationship among user satisfaction, decision making performance, and efficiency. Also, in the information systems success model presented by DeLone and McLean (1992), satisfaction is an effective variable on working efficiency or decision making level. DeLone and McLean (2003) conceptualized in their updated IS success model that user satisfaction is an antecedent of Net Benefits. In the study conducted by Igbaria and Tan (1997), the result showed that perceived user satisfaction has the strongest direct effect on individual impact. Iivari (2005) also found that user satisfaction is a strong predictor of individual impact (net benefits to individuals). This demonstrates the importance of examining user satisfaction in explaining net benefits to individual. Similarly, the intention to continue is predicted to have a positive relationship with net benefits. Thus, the following hypotheses are formulated:

H9: User satisfaction in e-training is positively related to net benefits.
H10: Continuance intention is positively related to net benefits.
Method

Sample and procedure

The population of this study was employees in organizations across different industries in Malaysia who had experience using e-training at their workplace. The primary target sample was employees working at multinational companies (MNC) in Malaysia as the majority of these firms had implemented e-training as a tool for the learning and development of their employees. The unit of analysis in this study was the individual who had experience using e-training in their organizations. A total of 194 responses were used in the final analysis.

Table 1. Profile of respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Frequency</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Between 18 to 29 years</td>
<td>129</td>
<td>66.5</td>
</tr>
<tr>
<td></td>
<td>Between 30 to 55 years</td>
<td>65</td>
<td>33.5</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>106</td>
<td>55.2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>86</td>
<td>44.8</td>
</tr>
<tr>
<td>Education</td>
<td>Certificate</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Diploma</td>
<td>50</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td>Bachelor Degree</td>
<td>123</td>
<td>63.4</td>
</tr>
<tr>
<td></td>
<td>Master Degree</td>
<td>17</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>PhD</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Nature of Job</td>
<td>Engineering</td>
<td>111</td>
<td>57.5</td>
</tr>
<tr>
<td></td>
<td>Information Technology</td>
<td>17</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Human Resource</td>
<td>16</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>Production</td>
<td>9</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>Sales/Marketing</td>
<td>7</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>Finance/Accounting</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>30</td>
<td>15.5</td>
</tr>
<tr>
<td>Number of years using the</td>
<td>1 – 3 years</td>
<td>11</td>
<td>5.7</td>
</tr>
<tr>
<td>Internet</td>
<td>4 – 6 years</td>
<td>22</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>7 – 9 years</td>
<td>52</td>
<td>26.8</td>
</tr>
<tr>
<td></td>
<td>More than 9 years</td>
<td>109</td>
<td>56.2</td>
</tr>
<tr>
<td>Total hours spent using the</td>
<td>1 – 3 hours</td>
<td>83</td>
<td>42.8</td>
</tr>
<tr>
<td>E-training system in a week</td>
<td>4 – 6 hours</td>
<td>42</td>
<td>21.7</td>
</tr>
<tr>
<td></td>
<td>7 – 9 hours</td>
<td>34</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>More than 9 hours</td>
<td>35</td>
<td>18.0</td>
</tr>
<tr>
<td>Number of Employees</td>
<td>Less than 500 employees</td>
<td>12</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>501 to 1000 employees</td>
<td>8</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>1001 to 2000 employees</td>
<td>16</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>More than 2000 employees</td>
<td>153</td>
<td>81</td>
</tr>
</tbody>
</table>
The study was conducted in a non-contrived setting, thus it did not disrupt the normal routine work of the organization. The method used to gather information in this research was through questionnaires composed of measures taken directly or adapted from previous studies. The research was a cross-sectional study. The primary data collection method was used whereby a survey was conducted using questionnaires. The inclusion criterion of sample selection is those employees who had prior experience in using e-training in the workplace.

Instrumentation

The constructs, concept definitions, items and sources of the instruments are depicted in Table 2.

Table 2. Conceptual construct definitions, items and sources

<table>
<thead>
<tr>
<th>Construct</th>
<th>Conceptual Definition</th>
<th>Measures</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-efficacy</td>
<td>Individual’s judgment of their capabilities to use computers in diverse situations</td>
<td>CSE1</td>
<td>Compeau &amp; Higgins (1995); Lim et al. (2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CSE2</td>
<td></td>
</tr>
<tr>
<td>Motivation to learn</td>
<td>Specific desire of the trainee to learn the content of the training course</td>
<td>MOT1</td>
<td>Baldwin, Magiuka &amp; Loher, 1991; Martocchio &amp; Dulebohn, 1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MOT2</td>
<td></td>
</tr>
<tr>
<td>Ease of Use</td>
<td>Degree to which a potential adopter considers use of the target system to be relatively free of effort</td>
<td>EOU1</td>
<td>Davis (1989); Lee &amp; Lee (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EOU2</td>
<td></td>
</tr>
<tr>
<td>Content of Training</td>
<td>Is concerned with such issues as accuracy, timeliness, completeness, relevance and consistency of the information generated by an information system</td>
<td>COT1</td>
<td>Bailey &amp; Pearson (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COT2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>COT3</td>
<td></td>
</tr>
<tr>
<td>Management Support</td>
<td>The aspect of work environment that reflects the extent to which supervisors and managers encourage on-the-job learning, innovation, and skill acquisition and provide recognition to employees in support of these activities</td>
<td>MS1</td>
<td>Tracey et al. (1995, 2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MS2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MS3</td>
<td></td>
</tr>
<tr>
<td>Organization Support</td>
<td>Policies, procedures, and practices that demonstrate the importance of training and development efforts, such as reward systems and resources to acquire and apply learned skills</td>
<td>OS1</td>
<td>Tracey et al. (1995, 2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OS2</td>
<td></td>
</tr>
<tr>
<td>User Satisfaction</td>
<td>Affective state related to the information system usage</td>
<td>SAT1</td>
<td>Spreng et al. (1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAT2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAT3</td>
<td></td>
</tr>
<tr>
<td>Intention</td>
<td>Intention to use the e-training system</td>
<td>INT1</td>
<td>Davis et al. (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INT2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INT3</td>
<td></td>
</tr>
<tr>
<td>Net Benefit</td>
<td>Positive effect from the use of the system</td>
<td>NET1</td>
<td>Torkzadeh &amp; Doll (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NET2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NET3</td>
<td></td>
</tr>
</tbody>
</table>

Result

Measures reliability and validity

Confirmatory factor analysis (CFA) was used to examine the reliability and validity of the measures adopted from the literature. The results are presented in Table 3. To measure the reliability of the measures we used the inter-item consistency reliability value of Cronbach’s alpha. The values range from 0.756 to 0.954 which were above the threshold of 0.7 as suggested by Nunnally and Bernstein (1994). Convergent validity that reflects the degree to which multiple attempts to measure the same concept is in agreement, was also examined. As suggested by Hair et al. (1998), factor loadings, composite reliability, and average variance extracted were utilized to assess convergent validity. The loadings for all items exceeded the recommended value of 0.6 (Chin et al., 1997). Composite reliability values, which depict the degree to which the construct indicators indicate the latent, construct range from 0.707 to
0.874 which exceeded the recommended value of 0.7 (Hair et al., 2010) except for self-efficacy and management support dimension which was only 0.610 and 0.634. The average variance extracted, which reflect the overall amount of variance in the indicators accounted for by the latent construct, were in the range of 0.610 and 0.874 which exceeded the recommended value of 0.5 (Hair et al., 2010). In addition, discriminant validity is examined by comparing the squared correlations between constructs and variance extracted for a construct (Fornell & Larcker, 1981). As shown in Table 4, the squared correlations for each construct is less than the average variance extracted by the indicators measuring that construct indicating adequate discriminant validity. In total, the measurement model demonstrated adequate reliability, convergent validity and discriminant validity.

Table 3. Result of CFA for measurement model

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
<th>Internal reliability</th>
<th>Factor loading</th>
<th>Composite reliability&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Average variance extracted&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-efficacy</td>
<td>CSE1</td>
<td>0.756</td>
<td>0.816</td>
<td>0.758</td>
<td>0.610</td>
</tr>
<tr>
<td></td>
<td>CSE2</td>
<td>0.745</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation to Learn</td>
<td>MOT1</td>
<td>0.828</td>
<td>0.857</td>
<td>0.828</td>
<td>0.707</td>
</tr>
<tr>
<td></td>
<td>MOT2</td>
<td>0.824</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of Use</td>
<td>EOU1</td>
<td>0.868</td>
<td>0.861</td>
<td>0.869</td>
<td>0.769</td>
</tr>
<tr>
<td></td>
<td>EOU2</td>
<td>0.893</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content of Training</td>
<td>COT1</td>
<td>0.878</td>
<td>0.907</td>
<td>0.885</td>
<td>0.721</td>
</tr>
<tr>
<td></td>
<td>COT2</td>
<td>0.873</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COT3</td>
<td>0.761</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Support</td>
<td>MS1</td>
<td>0.839</td>
<td>0.755</td>
<td>0.839</td>
<td>0.634</td>
</tr>
<tr>
<td></td>
<td>MS2</td>
<td>0.811</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS3</td>
<td>0.822</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization Support</td>
<td>OS1</td>
<td>0.879</td>
<td>0.852</td>
<td>0.886</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>OS2</td>
<td>0.931</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User satisfaction</td>
<td>SAT1</td>
<td>0.942</td>
<td>0.893</td>
<td>0.942</td>
<td>0.845</td>
</tr>
<tr>
<td></td>
<td>SAT2</td>
<td>0.944</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SAT3</td>
<td>0.920</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention</td>
<td>INT1</td>
<td>0.954</td>
<td>0.930</td>
<td>0.954</td>
<td>0.874</td>
</tr>
<tr>
<td></td>
<td>INT2</td>
<td>0.954</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INT3</td>
<td>0.921</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Benefit</td>
<td>NET1</td>
<td>0.945</td>
<td>0.908</td>
<td>0.945</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td>NET2</td>
<td>0.949</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NET3</td>
<td>0.911</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: <sup>a</sup> Composite reliability = (square of the summation of the factor loadings) / { (square of the summation of the factor loadings) + (square of the summation of the error variances) }

<sup>b</sup> Composite reliability = (summation of the square of the factor loadings) / { ( summation of the square of the factor loadings) + (summation of the error variances) }

Table 4. Discriminant validity of constructs

<table>
<thead>
<tr>
<th>Constructs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>(0.71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>0.03</td>
<td>(0.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of Use</td>
<td>0.25</td>
<td>0.01</td>
<td>(0.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content of Training</td>
<td>0.26</td>
<td>0.05</td>
<td>0.38</td>
<td>(0.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management Support</td>
<td>0.12</td>
<td>0.05</td>
<td>0.19</td>
<td>0.22</td>
<td>(0.63)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization Support</td>
<td>0.14</td>
<td>0.01</td>
<td>0.27</td>
<td>0.24</td>
<td>0.24</td>
<td>(0.79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Satisfaction</td>
<td>0.31</td>
<td>0.02</td>
<td>0.28</td>
<td>0.22</td>
<td>0.26</td>
<td>0.30</td>
<td>(0.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention</td>
<td>0.41</td>
<td>0.01</td>
<td>0.42</td>
<td>0.37</td>
<td>0.31</td>
<td>0.33</td>
<td>0.43</td>
<td>(0.87)</td>
<td></td>
</tr>
<tr>
<td>Net Benefit</td>
<td>0.29</td>
<td>0.02</td>
<td>0.26</td>
<td>0.24</td>
<td>0.22</td>
<td>0.24</td>
<td>0.28</td>
<td>0.52</td>
<td>(0.79)</td>
</tr>
</tbody>
</table>

Note: Figures in parentheses represent the average variance extracted while the other entries represent the squared correlations.
As shown in Table 5, it can be concluded that the full structural model meets all the goodness of fit indices.

### Table 5. Goodness of fit indices

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>Study</th>
<th>Recommended values</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>351.229</td>
<td>$\leq 3.00$</td>
<td>Bagozzi &amp; Yi (1988)</td>
</tr>
<tr>
<td>$\chi^2$/df</td>
<td>1.713</td>
<td>$\leq 3.00$</td>
<td>Bagozzi &amp; Yi (1988)</td>
</tr>
<tr>
<td>GFI</td>
<td>0.852</td>
<td>$\geq 0.90$</td>
<td>Chau &amp; Hu (2001)</td>
</tr>
<tr>
<td>AGFI</td>
<td>0.800</td>
<td>$\geq 0.80$</td>
<td>Chau &amp; Hu (2001)</td>
</tr>
<tr>
<td>CFI</td>
<td>0.953</td>
<td>$\geq 0.90$</td>
<td>Bagozzi &amp; Yi (1988)</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.066</td>
<td>$\leq 0.08$</td>
<td>Brown and Cudeck (1993)</td>
</tr>
<tr>
<td>NNFI (TLI)</td>
<td>0.943</td>
<td>$\geq 0.90$</td>
<td>Bagozzi &amp; Yi (1988)</td>
</tr>
</tbody>
</table>

#### Structural model

The structural model was estimated using the maximum likelihood estimate (MLE) offered by AMOS software version 16. We added an extra path from EOU to INT in the full structural model. The reason for doing this is that in structural equation modeling it is assumed that a full mediation is in place, whereas in this model there is a direct effect from EOU to INT and it does not go through SAT. As presented in Table 5, the test of the overall model fit yielded in a $\chi^2 = 351.229$ with 205 degrees of freedom and a $p$-value of less than 0.001. All the fit indices were above the recommended values except for the Goodness-of-fit index (GFI) which was slightly below the 0.90 cutoff. As shown in Figure 5, the explanatory power of the model ranges from as low as 0.57 for the effect on satisfaction. For the outcome variables the $R^2$ was 0.58 for net benefit and 0.69 for intention to continue use. Motivation to learn, management support and organization support were significant predictors of user satisfaction. Contrary to what was hypothesized, self-efficacy, ease of use and content of training were not significant. User satisfaction was also found not to influence net benefit but was a strong predictor of intention to continue usage as hypothesized.

![Modified structural model](image)

**Figure 5.** Modified structural model

#### Discussion

Hypothesis 1 which posited that there is a positive relationship between computer self-efficacy and user satisfaction was not supported, indicating that computer self-efficacy did not have an impact on user satisfaction in e-training.
Even though the result does not confirm earlier studies, there was an explanation to this phenomenon. Based on the respondent profile, the average hours spent on e-training per month was 4.23 hours. As the users gained experience in using the e-training system, the concern on the level of computer self-efficacy may no longer be an issue, as such; it did not impact the user satisfaction level significantly.

Hypothesis 2 which conjectured that there is a positive relationship between motivation to learn and user satisfaction was supported. This result was consistent with a few findings discovered in prior research by scholars. For example, Long et al. (2008) claimed that trainees’ motivation to learn was linked to their intentions regarding the amount of attention effort they intend to expend during the course. When the employees have high motivation to learn, they will exhibit greater interest in the e-training courses, and therefore, experience higher satisfaction.

Hypotheses 3 and 4 which stated that there is a positive relationship between ease of use with user satisfaction and also continuance intention showed mixed results. Surprisingly, ease of use was found to have a non-significant effect on user satisfaction, which contradicts earlier study by Shee and Wang (2008). On the other hand, ease of use was found to have a positive effect on continuance intention. A well-designed and user-friendly e-training system is therefore an important factor in increasing the intention to continue using the system. Employees will enjoy the learning process and continue to utilize the system when they can navigate through the e-training system easily to get the information they need, and when their interaction with the e-training system is clear and understandable.

As for hypothesis 5 which posited that there is a positive relationship between content of training and user satisfaction, it was not supported. There were a few possible reasons causing the insignificant relationship in this study. A plausible explanation for this perhaps lies on the respondents’ profile. From the profile, more than half of the respondents worked in American-based MNCs. These firms have implemented e-training for many years in providing technical, legal, and ergonomics trainings as well as company policies to their employees. The contents of training therefore were partially directly related to their work whereas the others were general trainings that were made compulsory but did not have direct applications to their real work.

Hypothesis 6 which stated that there is a positive relationship between management support and user satisfaction was supported. In today’s business environment, companies realize the importance of continuous learning of their employees to have a competitive edge over their rivals. Top management therefore has to provide support to their employees to keep on improving their knowledge and skills in the form of matching their staff’s needs for professional development via e-training, as well as provide recognition to employees who apply the knowledge and skills acquired in their jobs.

Hypothesis 7 which predicted that there is a positive relationship between organizational support and user satisfaction was also supported. As mentioned above, organizations today place great emphasis on continuous learning and improvement in order to compete successfully in the fast paced industry. With appropriate organizational support, employees will be motivated to exert greater effort in learning process, and also feel more satisfied with the e-training system.

Hypothesis 8 which conjectured that there is a positive relationship between user satisfaction and continuance intention was also supported indicating that the higher the level of user satisfaction in using e-training, higher level of effectiveness was observed. The finding attested previous research which found that individuals who are more satisfied with their experiences are more likely to enroll in future e-learning courses (Hayashi et al., 2004; Wang, 2003).

Hypothesis 9 which posited that there is a positive relationship between user satisfaction and net benefits however, was not supported. This is somewhat surprising given that perceived user satisfaction was found to be a strong predictor of individual impact (net benefits to individuals) (Igbaria and Tan, 1997; Iivari, 2005). Presumably, in the case of this study, even though the e-training courses were found satisfactory, employees may need some time to get accustomed to the new knowledge and skills acquired.

Finally, hypothesis 10 which predicted that there is a positive relationship between continuance intention and net benefits was supported. Generally, when employees’ expectations are met, they will exhibit more positive attitude towards the e-training system, and leads to continuance intention. Those who continue to use e-training system often
acknowledge the benefits that they have gained from the courses especially in terms of acquiring new knowledge and skills that will eventually improve their performance and productivity.

This study is not without limitations. Firstly, the sample is limited to the respondents working in MNCs in the Northern Region of Malaysia, with the focus of manufacturing companies in Penang. Therefore, the validity of the findings cannot be generalized to organizations across different industries in the whole of Malaysia. Secondly, this study only draws on data collected on cross-sectional basis. This limits the inferences with regards to causality between variables. Future research can also consider conducting a longitudinal study as it will provide a more meaningful and robust result.

Conclusion

In summary, this study sheds light into the important factors to be considered by organizations which have already adopted the e-training system or those considering embarking on this trend. In most cases, the focus of the e-training system is on the contents of the training as well as the information to be included; however, the ease of use of the system is often taken for granted. Given the findings that indicate the importance of ease of use, organizations should consider ways to simplify the navigation steps, provide clear instructions, and make the e-training system interface more attractive, interactive and easy to understand. The significance of both management support and organizational support in predicting user satisfaction draws attention to the importance of support from organization and that managers should take a more active role in ensuring the success of the e-training system.

This study offers valuable insights to the top management and IT managers pertaining to factors impacting e-training effectiveness. While organizations are well aware of the benefits that e-training brings, namely cost effectiveness, flexibility, convenience, consistency of contents across organizations; many do not recognize the critical factors that influence the success of the implementation. Potential e-learning benefits gained by the employees will help to ensure that organization achieves their return of investments.

References


Supporting Teachers in Designing CSCL Activities: A Case Study of Principle-based Pedagogical Patterns in Networked Second Language Classrooms

Yun Wen, Chee-Kit Looi and Wenli Chen
Learning Sciences Lab, National Institute of Education, Nanyang Technological University 1 Nanyang Walk, Singapore 637616 // yun.wen@nie.edu.sg // cheekit.looi@nie.edu.sg

(Submitted May 26, 2010; Revised November 20, 2011; Accepted December 11, 2011)

ABSTRACT
This paper proposes the identification and use of principle-based pedagogical patterns to help teachers to translate design principles into actionable teaching activities, and to scaffold student learning with sufficient flexibility and creativity. A set of pedagogical patterns for networked Second language (L2) learning, categorized and undergirded by rapid collaborative knowledge improvement (RCKI) principles, has been extracted and articulated in our school-based research. The patterns are distilled from a series of over 60 Chinese language lessons conducted in real classrooms which are co-designed by teachers and researchers over a 3-year period in Singapore schools. Each principle-based pedagogical pattern is presented succinctly in a simple pattern language. Two classroom scenarios of collaborative activities are elucidated, embodying some of the patterns to provide a more grounded feel of how they are enacted in practice. RCKI principle-based pedagogical patterns inform the design of collaborative language learning activities in networked classrooms. Preliminary findings reveal that principle-based pedagogical patterns provide teachers a workable way to design RCKI activities and enable teachers to adopt, internalize and adapt them in classroom use.

Keywords
Principle-based pedagogical patterns, RCKI, L2 learning, Computer-supported Collaborative Learning (CSCL)

Introduction

Spurred by rapid developments in information and communication technologies (ICT), diverse tools are now being used in a rich variety of ways to support teaching and learning. Goodyear (2005) uses the term networked learning broadly to include e-learning, web-based learning, online learning, and computer-supported learning. He emphasizes that “networked learning is not necessarily distance learning. It may well involve some face-to-face (FTF) interactions, whether or not these are supported by the use of ICT. The important element is connectedness” (2005, p. 83). In authentic classroom settings, teachers can harness such connectivity to increase opportunities for social interaction, as well as provide resources to support learning. The more dynamic and complex networked learning environment requires taking advantages of new opportunities and working with emergent rather than fixed goals (Sawyer, 2004).

Teachers face numerous problems in orchestrating networked classrooms, especially when dealing with complex pedagogies such as collaborative learning (Dimitriadis, 2010). Scripts are a popular means of guiding the design of computer-supported collaborative learning (CSCL) activities and supporting teachers in classroom orchestration by structuring the flow of the collaborative activity (Dillenbourg & Hong, 2008; Dillenbourg & Jermann, 2010). Dimitriadis (2010) attributes pedagogical patterns as a kind of mediating artifacts in a CSCL environment. Pedagogical patterns that are developed from analyzing expert knowledge of the practices of teaching and learning (Pedagogical Pattern Project, n.d.) can provide a more concrete and understandable form of knowledge for teachers instead. They can serve as a powerful way for sharing good practices (Dimitriadis, Goodyear & Retalis, 2009; Fincher, 1999). A growing number of studies have pointed out the potential of the pedagogical pattern approach for teacher professional development (DeBarger, et al, 2010; Law, Laurillard, & Lee, 2011; Prieto, Villagrá-Sobrino, & Dimitriadis, 2011).

However, pedagogical patterns that captured good practices in traditional teaching processes cannot be applied directly to CSCL activities in networked classrooms. Traditional pedagogical patterns have fixed goals and comparatively fixed routines. CSCL demands and enables shift the focus of education from learning as knowledge acquisition to learning as building shared meaning, enculturation into social practices and participation in valued activities situated within a community of practice (Roschelle et al., 2011). In such CSCL environments, teachers work with more emergent goal structures and opportunism in knowledge work (Zhang & Scardamalia, 2007). But,
exposing teachers to good use of these procedural-based pedagogical patterns without sufficient flexibility and theoretical support may not help them to enact them effectively in more unpredictable authentic classroom practices (Prieto, Villagrá-Sobrino, & Dimitriadis, 2011).

Zhang and his colleagues (2007, 2011) propose that pedagogical innovation requires the creative engagement of teachers to continually improve classroom designs and practices based on principle- rather than procedure-based approaches to teaching. However, drawing on our practical experiences, applying abstract principles of teaching and learning to real teaching practices is always a challenge for teachers, especially when they try to adopt new innovative teaching practices (Wen, Looi, & Chen, 2011). To optimize flexibility and accessibility of scripting, we propose the use of principle-based pedagogical patterns. These patterns are distinguished from procedure-based pedagogical patterns which are highly scripted and depend on routine expertise. They are intended to serve as flexible scaffolding for teachers to adopt or adapt in creating learner-centered activities in the classrooms. In this paper, we extract and articulate a set of principle-based pedagogical patterns for networked Second language (L2) learning, categorized and undergirded by Rapid Collaborative Knowledge Improvement (RCKI) principles.

The domain of learning Chinese language as L2 is selected, because although there are a growing number of studies concerned with pedagogical design in networked classroom practices, few are concerned with the disciplinary perspective of language learning. The major reason might be that the field of language learning has long been dominated by the traditional cognitive perspective in which individual internalization of mental processes and the development of grammatical competence (Firth & Wagner, 2007) are too much emphasized. It runs counter to sociocultural perspectives underpinning CSCL studies. Collaborative learning has long been widely used in language learning, whereas the existing collaborative learning studies on language learning focus too much on language itself, but without sufficient consideration of its function as a tool for collective thinking for the pursuit of joint intellectual activity (Mercer, 2004).

With the sociocultural perspective being more pervasive in the field of L2 learning, language learning is viewed as a semiotic process attributable to participation in social activities (Block, 2003; Lantolf & Thorne, 2006). Language learners are believed to be able to be able to enhance their linguistic knowledge and skills in well-designed collaborative learning through generating, sharing and improving their conceptual artifacts by interactional moves. Although a growing number of researchers work on developing collaborative technological environments to provide explicit scaffoldings for language learners, as well as visualization of and feedback on group work process (e.g., Reimann, et al., 2010; Yeh, Lo, & Huang, 2011), these studies focus on technology design instead of CSCL pedagogical design for real classroom environments. The pedagogical development in language teaching lags behind the development and use of ICT.

The context of this work is the introduction of a collaborative tool called GroupScribbles (GS) into primary and secondary school Chinese language classrooms in Singapore through a process of design research. Working closely with teachers, researchers have co-designed GS-based collaborative activities in over 60 Chinese language lessons. To guide the design of the lessons, the notion of RCKI was developed, and 9 design principles have been postulated and shared with the teachers. Thus, a principle-based approach was adopted in the GS lessons design for the first three years. Since the middle of 2010, by analyzing the various classroom lessons, we have identified and distilled principle-based pedagogical patterns and introduced them to new teachers in new rounds of school intervention. The present research is concerned with: 1) What are RCKI principle-based pedagogical patterns for L2 learning? 2) How these pedagogical patterns can be enacted as classroom practices by expert practitioners? 3) What is the effect of principle-based pedagogical patterns as a professional development approach to help teachers deploy technology-enabled collaborative practices? Descriptive case studies are used to address the questions.

**RCKI and its principles**

In our work, we have proposed the RCKI concept, which refers to the notion of democratizing participation and idea refinement in the context of live dynamic classroom settings, that is, FTF collaborative knowledge construction and improvement over the duration of a class session, and supported by certain technologies for lightweight instant interaction. RCKI seeks to harness the collective intelligence of groups to learn faster, envision new possibilities, and to reveal latent knowledge in a dynamic live setting (DiGiano, Tatar & Kireyev, 2006). When enacted in the classroom, RCKI takes the form of alternative ways to promote classroom interactions that enable students to co-construct knowledge and learn content skills. It is designed to address the constraints faced by classroom teachers.
when they are designing and implementing knowledge construction and improvement practices within the short duration of a classroom lesson ranging from say half an hour to one and a half hours. The notion of “rapid” is understood from 3 main aspects of a learning activity: 1) it is done within a limited time of participation; 2) it uses a lightweight form of expression; 3) it must enable the participants to have quick cycles of interaction. RCKI focuses on democratic knowledge sharing as well as cycles of individual and group knowledge enhancement.

The concept of RCKI seems suitable for guiding language learning, especially L2 learning. L2 teachers typically handle the low proficiency of students in the target language by focusing primarily on vocabulary and grammar, and hence approaches like ideas generation and expression are seldom adopted as learning activities (Scott, 1996; Stapa & Majid, 2009). When cognitive load is not just expended on language expression, the enthusiasm and capacity of L2 learners to engage in reading and writing in on-line based interaction could be enhanced (Wen, Chen, & Looi, 2010).

Based on more than two years of school-based research, we have articulated a series of RCKI principles (Looi, Chen, & Patton, 2010). Our research team first articulated 9 characteristic principles of RCKI, some of which are built upon the knowledge building theory and its principles proposed by Scardamalia (2002). In this paper, we categorize these 9 RCKI principles into different clusters in terms of the roles they played in fostering idea improvement. We believe categorizing these principles can enable us to better understand the feature of each principle as well as the interfaces between them, and thus provide a way to mine pedagogical patterns engendered by them.

**Cluster 1: Principles that can be realized by activity design that promotes students’ participation**

1.1 Spontaneous participation: designing for quick, lightweight interaction driven by students themselves;
1.2 Multimodal expression: accommodating different modes of expression for different students;
1.3 Volunteerism: let learners choose what piece of the activity they want to participate in.

**Cluster 2: Principles that rely on or seek to transform students’ attitudes**

2.1 Democratized knowledge: everybody participates and is a legitimate contributor to knowledge;
2.2 Epistemic agency: encouraging students to take responsibility for their own and one another’s learning.

**Cluster 3: Principles that focus on ideas**

3.1 Idea diversity: supporting the exploration of ideas and related/contrasting ideas, and encouraging cognitive diversity;
3.2 Improvable ideas: supporting the exploration of ideas and related/contrasting ideas, and encouraging idea improvement.

**Cluster 4: Principles that require or foster higher levels of collective cognition**

4.1 Higher-order thinking: encouraging skills like analysis, synthesis, evaluation, sorting, and categorizing;
4.2 Symmetric knowledge advancement: expertise is distributed, and advanced via mutual exchanges.

**Pedagogical patterns**

The term “pedagogical patterns” originated from the concept of design patterns. The word “pattern” as coined by the architect Christopher Alexander (1977, 1979) refers to the core solution to a recurrent problem in a concrete context. The notion of design patterns is also adopted in software design and development where it is used to capture and share aspects of the software engineering experiences, and is a way of representing successful models for system implementation (Goodyear, 2005).

In education, pedagogical patterns deal with the problem of how to teach (Fricke & Völter, 2000), by providing structures to encode, share and use knowledge among instructors (Goodyear, 2005). The idea of patterns for pedagogy as a means to present teaching experience is originally suggested by Lilly (1996) in the notion of “reusable pedagogical design patterns.” The features include: a) they should be repeatable and easy to adapt; b) they should be described in a way that allows them to be easily applied for different lessons by different instructors; and c) they do not have to be novel or original; rather, they should communicate proven solutions to common problems. “Pedagogical patterns play the same role in teaching and learning that design patterns play in software development” (Seffah & Gregono, 2002, p. 18). The terms pedagogical design patterns and pedagogical patterns are often used synonymously.
“Pedagogical patterns have a number of qualities which, in combination, give them potential to be an effective way of sharing experience in networked learning” (Goodyear et al., 2004). It is also pointed out by Goodyear that the approach of pedagogical patterns “can be seen as a way of bridging between teaching philosophy, values, theory, empirical evidence and experience (on the one hand) and the practical teaching context of design” (2005, p. 92). Most research studies focus on highlighting the effectiveness of pedagogical patterns for novice teachers. In the context of networked learning, we would like to emphasize that it cannot only effectively support novice teachers in teaching, but also can assist the experienced teacher to better understand the philosophy of a new teaching environment such as that of networked classroom environment.

In the following section, we present the characteristics of a network classroom technology GS that is our vehicle for studying the RCKI principles and its corresponding pedagogical patterns in the context of L2 learning.

**Technology design**

In recent years, some researchers delved into the design of interactive technologies to support students’ active classroom participation by harnessing the collective intelligence inherent in the classroom. One of the technologies is GS, co-developed by SRI International and National Institute of Education of Singapore, which enables collaborative generation, collection and aggregation of ideas through a shared space based upon individual effort and social sharing of notes in graphical and textual form (SRI International, 2006). We discuss RCKI in the context of our experiences with GS that designed to be lightweight, flexible for collaboration, content independent and easy-to-improvise by teachers. It attempts to maximize the power of ink, improvisation, and interactive engagement, so that teachers can improvise different patterns of collaborative activities for students without the need for additional software programming (Roschelle et al., 2007).

![Figure 1. A screenshot of a collaborative activity from the GS user interface](image)

We now describe the current version of GS 2.0 and its features. The GS interface consists of a multi-pane window. The default configuration consists of 2 panes: a lower pane and an upper pane, but the user can slide in more panes as desired. The lower pane is usually the private board, or the user’s personal work area, with a virtual pad of fresh “Scribble Sheets” or notes on which the user can draw or type (Figure 1). The upper pane is usually a public board or group board, into which users can post their Scribble Sheets, position them relative to other’s Scribble Sheets and take items back to the private board for amendments or elaboration. When any Scribble Sheet is posted, moved or updated, others can see the effect almost immediately. On each pane, there is a drop-down menu to allow users to switch to other boards. Students post anonymously so as to freely express their ideas. Figure 1 shows a GS activity on fractions in which students in a group participate by posting various possible representations of a fraction like 1/4.

**Framework for research design**

Figure 2 depicts the framework for our research process. Our RCKI principle-based pedagogical patterns are distilled and induced from around 60 GS lessons, based on first three years of school-based research in which we support teachers from three schools that co-designed the patterns with us and enacted them in real classroom settings. This
The paper focuses on the first three stages of the research process framework that are concerned with the process of pedagogical pattern generation and the preliminary finding of the effect of patterns application.

In the following section, we further articulate the corresponding pedagogical patterns engendered by these principles. We then provide two classroom scenarios in the GS-based Chinese language learning classrooms. These two cases serve these purposes: It provides a further elucidation for some of the patterns and their interrelationships, and it helps address the question how these pedagogical patterns can be enacted in authentic classroom good practices by expert practitioners.

The pedagogical patterns generated by RCKI

Enumerating pedagogical patterns in a sensible way requires the adoption of a pattern language that facilitates the description of patterns, the creation of new patterns, and the editing and improvement of existing patterns. Goodyear and his colleagues advance some tentative proposals about an equivalent pattern language for designing networked learning. But they have realized that forming a pattern language for networked learning involves painstaking, iterative work (Goodyear, 2004). Thus in building our pedagogical patterns in the context of RCKI-enabled second language learning, we look for a compatible pattern language which we can adopt or modify.

Towards the task of categorizing pedagogical patterns, Bennedsen and Eriksen (2006) set out a universal pedagogical pattern categorization based on teaching values and activities, and indicate that it could be more sustainable than arbitrary categorization implied by pedagogical pattern language themes. Likewise, we delve into building and organizing pedagogical patterns with respect to RCKI principles. The ultimate goal of any of the proposed principles could be to solve a core problem as well. Different patterns can serve as a number of building components that can be creatively combined into new solutions (Prieto, Villagra-Sobrino, & Dimitriadis, 2011). Building pedagogical patterns based on the principles can not only depict the characteristics of RCKI-based pedagogies, but also reveal the hierarchies and the relationships amongst the patterns. Principle-based pedagogical patterns hence can provide a deeper understanding of pedagogical patterns to be enacted in practice to ensure their success in achieving learning efficacies.

| PATTERN NAME | Problem Description | Solution outline, consequences, drawbacks | An example of the pattern which is enacted in practice |

Figure 3. Sample of pedagogical pattern form
We mainly adopt the pattern language proposed by Fricke and Völter (2000). Their work of pedagogical patterns on *seminars* is noted as a very successful pedagogical pattern language (Pedagogical Pattern Project, n.d.), which is similar to the earlier one created by Alexander. Our pattern includes six sections (see Figure 3). The first section in the middle top is the pattern name. The problem description is the core section of the pattern and is highlighted in bold font in the second section. The third section used asterisks (*) to indicate the difficulty of the pattern implementation. Each pattern is marked with one to five asterisks (from the easiest to the most difficult to enact). The fourth section (below the ***) outlines the solution to the described problem as well as the consequences, limitations, and drawbacks. The last section provides the examples on implementing the pattern. As in Fricke and Völter’s patterns, references to other patterns in this language are highlighted in CAPITAL LETTERS.

**RCKI principle-based pedagogical patterns**

1. **SPONTANEOUS PARTICIPATION**

   Students need to be given opportunities in a conducive environment to articulate their opinions, questions, answers and comments spontaneously in a class lesson or activity. Students can view each other contributions spontaneously as well.

   *

   Quick and lightweight tasks may provide better opportunities for motivating each student to contribute their own ideas. Students realize that whatever they post or contribute can be seen by the whole class, so that they should be responsible for their postings.

   Examples from words/idioms/proverbs learning: The instructor can read out the words to the students, and immediately students try to write them down and post them into the public board. The instructor or peers can give instantaneous feedback. Students can help others to correct writing mistakes, and learn language skills from one another by imitation.

   Students can also be involved in contributing relevant or associated words around a central word or theme provided by the instructor. They write down words or concepts associated with the central word or theme much like creating a concept map.

2. **MULTIMODAL EXPRESSION**

   It is often challenging for L2 learners or novice language learners to use or write characters or words to express their own ideas because they do possess sufficient language capability.

   *

   With technology that supports multimodal forms of expression, the students can have an opportunity to choose alternate ways to express themselves. Those students with low language proficiency can still contribute to group work. Students, who are good at speaking, may know how to pronounce the word but without knowing how to write it. In Chinese language learning, they are able to type out the Chinese words or directly write in Hanyu Pinyin. As long as the content matches the topic, expressing ideas by drawing is also acceptable.

   For example, after reading a piece of text, the instructor can ask students to generate a new ending for the story based on their own interpretation of the original text. Students are also encouraged to draw animated pictures and then provide a proper text description. This can encourage students to be engaged in the task as well as motivate their peer groups to review and give comments for their work.

3. **VOLUNTEERISM**

   Students may feel less motivated and show little initiative if they are not interested in the task assigned to them. A task is decomposed into subtasks with which students can choose specific tasks to do, thus building up
their self-esteem to be partners in collaborative work with others.

**

The instructor allows students to choose the piece of the activity he or she wants to contribute towards. In intra-group work, the approach of having students choose sub-task(s) themselves can provide opportunities for giving full play to their specific knowledge and skills. Volunteerism becomes quite natural for a mature group, in which all the students have known how to do collaborative work.

In the comprehension of a piece of text, the instructor poses comprehension questions or tasks in the public space. The instructor can also pose new vocabulary for the students to further explore. Students choose the question or task they are interested in, and work on them.

4. DEMOCRATIZING KNOWLEDGE

Everyone in a group has a right to and can contribute to the group’s cognition and to producing the group artifacts. The group task should not be dominated by one or two members.

**

The group task, including several sub-tasks, should be designed with different difficulty levels, and it is necessary to make sure that the task cannot be completed independently by one or two group members in a short time. Thus each student must do their part in completing the task, regardless of their second language proficiency.

In learning new vocabulary, the instructor designs a template, which is uploaded as the background organizer in the group space for the students to collaboratively complete their task of learning about the target vocabulary words. The subtasks may include explaining the meaning of the word, giving a few homophones or similar characters as the main character of the word, word formation, etc. After posting their own contributions, individual students need to evaluate and organize their information within the group. Different groups are in charge of different vocabulary words, so that all students can learn more words by doing a gallery walk of other group boards.

5. EPISTEMIC AGENCY

In some groups, some students, especially those students with weak L2 proficiency, face challenges. They may merely focus on finishing their own piece of task, and do not care about what others are doing. Students with high L2 proficiency may be contented with just completing own group work, but rather than consider what they can learn from other groups.

***

Spontaneous participation is the premise for the form of epistemic agency. A comfortable group environment ensures epistemic agency in group members can be cultivated successfully. Democratizing knowledge and epistemic agency are two interrelated concepts. Many opportunities for intra- and inter-groups interaction should be provided so that students may be encouraged to take responsibility not only for their own but also for one another’s learning.

An activity on collaborative writing which focuses on inter-group interaction is to have different groups follow a sequence to complete the different portions of a story. As several groups or a whole class is required to create a story together, they have to constantly change ideas or negotiate with others throughout the writing process.

6. IDEA DIVERSITY

Language learning, especially L2 learning, may be too concerned with learning linguistic knowledge, neglecting another aspect of using language, namely, the creative capacity. Students may be accustomed to absorbing linguistic knowledge from teachers. Thus their L2 learning may include such elements of rote learning. Students’ own thinking and emotions are seldom foregrounded in L2 writing and reading.
Students drawing on their own personal experiences, values or culture would generate diverse ideas for the same topic, and they can share them in the way in which they feel comfortable. It would be better to design tasks requiring diversity, thereby providing more students with opportunities to contribute their ideas. The task/question should be open-end. If necessary, the teacher can require students to give more than one proposal or answer for each task/question.

When delivering creative writing lessons, the instructor can first assign a writing topic (or a series of cartoon pictures for essay writing) to students, and then ask the students to share their ideas in developing their story plot. After completing their own group task, the students can look at other group boards to appropriate the ideas from other groups.

7. IMPROVABLE IDEAS

When students are asked to give comments for each other language constructions, they typically correct each others’ writing or grammatical mistakes, but do not focus or do not know how give suggestions to improve others’ ideas.

Improvable ideas often appear together with idea diversity, while it requires more of a cognitive ability to comment or build on someone’s ideas. Considering students’ diverse abilities in the same class, more opportunities should be provided for intergroup interaction. In addition to the requirement on the difficulty and openness of the tasks/questions, the students should have sufficient time to discuss and negotiate with one another.

To better understand a word/idiom/proverb, students in a group can create a scenario for using the target vocabulary words. Within the group, each student has an equal opportunity to contribute their ideas and build a deeper understanding of the words through discussing and negotiating with others.

To help students better understand the text, the instructor can design a template to assist students with developing a deeper understanding. With mutual scaffolding, group students improve their understanding of the text through discussion and negotiation.

8. HIGHER ORDER THINKING

Students are always good at knowledge sharing, but may not have the need to do higher-order thinking. Limited by their linguistic knowledge, they struggle to achieve skills like synthesis, evaluation etc.

Provide students time to organize their postings, giving full play to the characteristics of “lightweight.” Let students synthesize their understanding and share with peer students. Offer metacognitive scaffolds to assist students’ higher-order thinking.

When doing reading comprehension, the students are required to outline the meaning of each paragraph in groups, and then distill the core meaning of the reading text after peer negotiation.

9. SYMMETRIC KNOWLEDGE ADVANCEMENT

Some groups are always able to complete their tasks quickly while some groups are often unable to finish their work in the given time. The same problem occurs in individual students when they are involved in group work.

****
Democratizing knowledge and symmetric knowledge advancement are closely related with each other. Positive mutual interdependence is necessary for the achievement of symmetric knowledge advancement. It requires much interaction within and across groups.

For reading comprehension, a version of reciprocal teaching is used: A group poses questions for each group, and vice versa. The group will also evaluate the answers provided by the other group and provide formative feedback.

Classroom scenarios using principle-based pedagogical patterns

A scenario of Chinese characters learning

We have worked with three schools in the use of GS for CL2 in 2008 and 2009. From these GS lessons, we pick one scenario of Chinese words/characters learning based to elaborate on the 2 patterns, namely, volunteerism and democratized knowledge. In this scenario, 40 students in a primary school class were classified into 10 fixed groups. Their teacher defined 10 new Chinese words, and asked each group to choose one of them to work on. The purpose of this GS activity on learning new words/characters was to enhance students’ awareness of the character components, and to help students familiarize with the usage of the words by composing sentences that make use of the words. Two kinds of templates with different difficulty levels were designed and uploaded by the teacher as the background organizer of GS group board. Figure 4 and Figure 5 show two GS group boards using the graphic organizer.

In this lesson, the students first read a Chinese text in class. The teacher identified a list of new vocabulary, namely, characters or phrases for the groups of 4 students each to study further as a group. As shown in Figure 4, the group was first required to write down in the centre of the GS board the character/phrase with its Chinese phonetics (Han Yu Pin Yin). In this case it was a phrase “不理不睬” (meaning “to completely ignore”). The space surrounding the phrase was divided into 7 sections, each of which was dedicated for one task. Starting from the top right in clockwise order, the tasks were: 1) to explain the meaning of the phrase; 2) to give a few similar characters as the last character of the phrase; 3) to give a few homophones of the last character of the phrase; 4) to use the last character of the phrase to form words (word formation); 5) to read comments from other groups (similarly this group was required to give comments on other groups’ work); 6) to expand the phrase into a sentence, and 7) to collocate the phrase with other words (word collocations).
Figure 5. Another shared group board with organizer for Chinese vocabulary learning

The organizer shown in Figure 5 was simpler. In addition to writing down the target word “得罪” (meaning “offend”) in the centre of the board, the students in the group were only required to complete the tasks 1, 5, 6, and 7. In this way, the **volunteerism pattern** can be seen at work. The students in the group can select a task from a list given by the teacher which varies difficulty levels, by considering their group ability. The students also have an opportunity to select the sections they are interested in or are familiar with to collaboratively complete them.

Underlying the intent of this activity is also the **democratizing knowledge pattern**. Due to the subtasks with different difficulty levels and themes, all group members have an equal opportunity to contribute their individual knowledge and ideas. Students who are not good at Chinese language can post the meaning of the phrase/word with the help of a dictionary. Students with higher Chinese language capability can take the responsibility for writing the sentences. Each student with diverse ability language capability can contribute diverse homophones and similar characters for the given characters. Within the group, the task will not be dominated by one or two students. Rather, all the students can play their role on completing the task.

When all groups have done their work, the teacher orchestrated sharing between the groups at the class level. The teacher asked the students to visit other group boards to learn the words which were worked on by other groups, and to provide comments to help others refine their group products. Students with different Chinese language proficiency and groups with different ability levels will have an equal opportunity to receive and provide comments anonymously. In Figures 4 and 5, we can see the positive criticisms (as circled in the figures) from other groups.

This scenario of GS activity took place in the beginning stage of integrating GS with Chinese language learning in this class. Although the students were asked to give comments on other students’ artifacts, the majority of their comments are consisted of words of praise and encouragement, but they were not constructive enough (see Figure 4). In the following scenario of reading comprehension, we describe another scenario of GS activity in a secondary class. In this case, when the GS activity was conducted, the class students had been involved in our project for a whole year (one GS lesson per week). We have observed that the culture of RCKI in the class had already been fostered on a certain degree. In this sense, it provides a prerequisite for enabling the more challenging pedagogical patterns to be enacted and used effectively.

**A scenario of reading comprehension**

This GS activity in a scenario of reading comprehension embodies the patterns of **improvable ideas** and **symmetric knowledge advancement**. The context of this activity is a 70-minute lesson for a Secondary 2 (Grade 8) class. In this scenario, the teacher wanted the students to develop a good understanding of a given text by having them engaged in collective cognition. To enable the students to contribute more diverse ideas and to seek continuous improvement of their ideas, they were required to do both intra-group and inter-group work, thus realizing the knowledge advancement not only at the group level but also at the whole class level. The activity comprises mainly the following 5 stages.
1) **Reading the article**: At the start of the activity, the students were given 5 minutes to read the article from the textbook. The article depicted a wretched life story about a woman called **Jin Jie** who had never married and worked her entire life as a maid servant instead. The students in the class of 40 were divided into groups of 4 students each.

2) **Generating questions**: The groups were given 20 minutes to set 3 text comprehension questions with different difficulty levels onto their group board, based on their understandings of text. The default score for each question was assigned depending on the difficulty, while the sum scores of all the 3 questions must be 10.

3) **Addressing questions**: When all the groups finished creating their group questions, two adjacent groups (e.g., group 1 and group 2, and so on till group 9 and group 10) exchanged group boards and answered the questions given by their partner group.

4) **Judging answers**: Students went back to own group board, evaluated and graded the answers of their partner group.

5) **Explaining**: In the last 20 minutes, each group selected one question with the lowest score from the 3 questions set by their own group, and posted it onto the public board for the whole class (see Figure 6). The teacher would complete the lesson by leading a class discussion.

Like a peer version of reciprocal teaching (Palincsar & Brown, 1984), the formulating question strategy guides the students in a group to interact with the text in more sophisticated ways. As students had ownership in creating the questions and were involved in answering the questions of others, there was mutual interaction among the students in a group, and healthy competition between partner groups. In this sense, the pattern of **improvable ideas** was enacted to elicit a plethora of diverse ways of composing appropriate comprehension questions. The processes of idea improvement could be found in the episodes in which students were generating their group questions, negotiating their group answers, and interacting with the partner group students.

Let us explore the interactions of a particular group. Group 10 was constructing questions for a text. The group members, Phoebe and Lucy, possessed good Chinese language proficiencies. Their Chinese language proficiency differentiated the depth of discussion of group 10 from that of group 6. Phoebe and Lucy pushed each other to think about the question more deeply.

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Conversation/Action</th>
<th>Transcript/Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Phoebe</td>
<td>这是什么意思啊？你看。会觉得她无法梳理自己的愁绪？</td>
<td>Phoebe first asked Lucy what the sentence “cannot sort out her own melancholy” refers to in the article.</td>
</tr>
<tr>
<td>2-2</td>
<td>Lucy</td>
<td>就是她现在很多问题啊。她不懂要回家还是要留在这边。她怕有病。</td>
<td>Lucy told Phoebe “it means she (Jin Jie) has a lot of problems now. She does not know she should go home or stay in here. She is afraid of sickness.”</td>
</tr>
<tr>
<td>2-3</td>
<td>Lucy</td>
<td>很多问题。她自己老了。她不懂要不要回家，因为她很多同伴的遭遇。</td>
<td>“A lot of problems” Lucy emphasized. “She has been old. She does not know whether she should go back home, as she knows the misfortune of her friends (such as Fang Jie).” Lucy added</td>
</tr>
<tr>
<td>2-4</td>
<td>Phoebe</td>
<td>And then?</td>
<td>“And then?” Phoebe encouraged Lucy to continue.</td>
</tr>
<tr>
<td>2-5</td>
<td>Lucy</td>
<td>她不知道怎么做。</td>
<td>“She does not know how to do.”</td>
</tr>
<tr>
<td>2-6</td>
<td>Phoebe</td>
<td>可不可以说如果留住，她留在这边的话，她会像她朋友(芳姐)这样啦？</td>
<td>“Can I say once she stays in here, she will be as unfortunate as her friends?” Phoebe asked.</td>
</tr>
<tr>
<td>2-7</td>
<td>Lucy</td>
<td>她留在这边，她怕她没有钱。她怕她很多问题。</td>
<td>“If she stays in here, she fears that she has no money. She worries about lots of issues.” Lucy highlighted this was the key point.</td>
</tr>
<tr>
<td>2-8</td>
<td>Phoebe</td>
<td>Then. 如果她又回去呢？</td>
<td>“Then, what will happen, if she goes back home?” Phoebe continued to ask.</td>
</tr>
<tr>
<td>2-9</td>
<td>Lucy</td>
<td>如果她回去她又怕她的 parent，她的亲人不要管她。你看这边白眼嘛。</td>
<td>“If she goes back home, she would worry that her parents, and her relatives may ignore her. You can see here, white eyes.” Lucy answered.</td>
</tr>
</tbody>
</table>
| 2-10 | Phoebe| 为什么会白眼她(Fang Jie)？为什么                                                       | “Why do they (Fang Jie’s relatives) white eyes her
Lucy: Ah! Because they work in other places, when they go back home, they have been too old. Lucy was getting a little bit impatient.

Phoebe: "But, why do they (Fang Jie’s relatives) ‘white eyes’ her (Fang Jie)?" Phoebe insisted. She still could not get the interrelationship between the old and white eyes.

Lucy: Ok, teacher I want to ask you a question. Why does she get white eyes from others?

Phoebe: “Fang Jie” is literally translated as “white eyes” here. In English its exact meaning is “roll the eyes,” while in the following excerpts, we translate it as “get a stare.” As “stare” in English is often used as verb, its corresponding Chinese word “白眼” is used incorrectly in students’ discussion.

From the code 2-1 to 2-9, Lucy followed Phoebe’s question to interpret the sentence step by step. As they discuss further details, they use language inappropriately (code 2-10). As shown in episode 2, Lucy answered Phoebe’s question patiently (code 2-11), but Phoebe still thought Lucy did not answer the key of the question (code 2-13). Lucy had to ask the teacher for help. Like Lucy told Phoebe (code 2-11), the teacher mentioned to students to pay attention to the paragraph that “Fang Jie has been old,” but Phoebe still could not understand. She asked, “Is there anything wrong to go back home when she is old?” The teacher reminded her Fang Jie was not married so she had no child and nobody could take care of her. In this way, Phoebe seemed struck by this new awareness, and Lucy nodded her head many times in agreement as well.

They afterwards planned to post this question for the partner group to answer. While they were organizing and choosing words for constructing the question, Lucy corrected the use of “白眼” for her group members, and finally they posted their question “为什么芳姐会遭到亲人的白眼？（Why does she get white eyes from others?）.” At the last stage of the activity, when the teacher asked each group to post the hardest question onto the class public board (see Figure 6), group 10 posted this question and explained the answer in front of the class.

* “白眼” is literally translated as “white eyes” here. In English its exact meaning is “roll the eyes,” while in the following excerpts, we translate it as “get a stare.” As “stare” in English is often used as verb, its corresponding Chinese word “白眼” is used incorrectly in students’ discussion.

On the basis of the collective cognition through intra-group negotiation, inter-group discussion and the whole class level sharing, we distill the pattern of symmetric knowledge advancement being enacted. This lesson scenario shows that the success of symmetric knowledge advancement depends very much on the pattern of improvable ideas.
Effects of RCKI principle-based pedagogical patterns as a pedagogical development approach

Based on the work with the 3 schools, we have identified and distilled these principle-based pedagogical patterns. In working with new teachers or new schools, we would like to explore the use of these patterns by new teachers. In the middle of 2010, a new secondary school joined our project to adopt and use GS in their CL2 lessons. Before the teachers conducted formal GS lessons, we conducted 4 rounds of professional development in which RCKI principles and its related pedagogical patterns were introduced to our new Chinese language teacher, Carol. She had approximately 10 years of teaching experience, but she was new to the use of GS. While she had a strong desire to use ICT in teaching, she lacked the capability and the confidence on how to do it well. Like most Chinese language teachers, she mainly adopted the drill-and-practice approach in her teaching. In class, Carol and her students spent most of classroom time doing two things: deciphering linguistic forms and meanings at the word and sentence level, and understanding the context of a given text. Reading comprehension was usually accomplished through teacher-centered explanations.

After sharing with her a number of past lessons which incorporated the principle-based pedagogical patterns, we asked Carol to choose the patterns to adopt by herself, and helped her transform her lesson ideas into GS activities. Within half a year, Carol had co-designed with us and implemented 7 GS lessons (each per week) for a Secondary 1 (Grade 7) class. The patterns she used include multimodal expression pattern and idea diversity pattern for teaching writing, and spontaneous participation pattern for teaching Chinese new words and phrases. Carol designed and implemented concrete lessons creatively on the basis of these patterns. For instance, while addressing the aforementioned patterns of improvable ideas and symmetric knowledge advancement for teaching reading comprehension, Carol designed and carried out the activity which comprises 1) reading the article; 2) generating questions; 3) addressing questions; 4) judging the answers and 5) explaining.

![Figure 7. Public board for the whole class in Carol’s class](image)

The students were asked to read a text article entitled “Be prepared for danger in times of safety and security” within the given time. Similarly, each group was given 20 minutes to compose text comprehension questions. She asked each individual student to post his/her questions on the public board of each group within 8 minutes. Then she asked each group to select out the questions that deserve further discussion in the rest 12 minutes. Specifying individual work to be done within 8 minutes deliberately encourages each student to spontaneously and concurrently participate in the group discussion. In this sequence, every student was expected to represent his/her ideas visually to pave a path for the further improvement of his/her ideas. Group discussion otherwise is often dominated by students with comparatively higher Chinese language ability. After that, groups worked in pairs, and each group addressed questions and judged answers that were provided by its pair group. The design of the final step was slightly altered.
During the last 20 minutes, each group selected the best question and posted it onto the public board for the whole class sharing (see above Figure 6). Carol gave instructions that each group was to select not only the best question (instead of the question with the lowest score) but also its related answers to share (see Figure 7). The teacher had deemed it difficult for students at grade 9 to judge the quality of questions, but proposed that selecting the question in which they were interested and sharing their answers towards the question seemed more productive for the students. In this way, the teacher brought to the fore for class discussion students’ (potential) comprehension of the solution, evidenced through their concrete artifacts. Thus teacher’s role is more than coordinating the process of group work, but is to find coherence with the different students’ answers and observations (Pata, Sarapuu, Lehtinen, 2005).

In a semi-structured interview with Carol, she shared:

Initially I do not really understand how to go about designing the activity. After all, there is some technicality involved. … I have read the principle of using GS. That’s it. I was just reading. In the end I’d still have to implement them. I have only read and understood the principle theoretically. Between understanding and applying them in classroom, there is still some distance to cover.

Carol said that she became more confident of conducting the GS lesson using these principle-based pedagogical patterns. She mentioned that she derived a better understanding of collaborative activity design and RCKI, and she felt that she had internalized them.

**Discussion and conclusion**

The design and enactment of collaborative activities in a networked classroom environment is a very complex process that has to take into account a multitude of factors. One way to start is to expose teachers to best practices, either written or as depicted on videos. Patterns make pragmatic resources more accessible to practitioners (Mor et al., 2006) by encapsulating the knowledge from experts and experienced teachers in a form which is transferable and applicable to similar classroom situations. However, some studies have suggested that innovation of teaching practices are intertwined with teacher beliefs (Chen, Looi, & Chen, 2010; Jacobson, et al., 2010). If a teacher cannot understand the essence of design behind pedagogical patterns, s/he can only emulate a similar process of activity, but the efficacy of pedagogical patterns may not be fully unlocked. An approach based on starting from principles has been postulated to help teachers to internalize an innovative mindset (Zhang & Scardamalia, 2007). The principle-based approach can engage teachers in reflective interpretation and in making adaptive classroom decisions to orchestrate the classroom activities well. But it is always a challenge for teachers to comprehend such principles especially when they appear to be rather abstract or de-contextualized. Grounding teachers with principle-based pedagogical patterns, hence create a focus for the teachers to start with but to lead eventually for them to understand the principles behind the pedagogical patterns.

In this study we propose that principle-based pedagogical patterns, used in conjunction with classroom network technologies, have the potential to advance pedagogical knowledge of classroom practices and help teacher understand abstract principles and address major challenge to effective use of technologies. Towards L2 learning in GS-based networked environment, we articulate 9 pedagogical patterns categorized and undergirded by RCKI principles. They inform the pedagogical design of CSCL activities taking account of the characteristics and requirements of language learning in a networked classroom environment.

We share our preliminary findings concerning one teacher from a local secondary school who has done an uptake of the pedagogical patterns which were developed and identified from earlier work with other schools and teachers. Our approach to having the teacher understand and apply the principle-based pedagogical patterns is via professional development. Much time and effort is invested in such efforts and in co-designing lessons with our teacher. Future research studies should be concerned with how often and reliably our teachers can integrate these patterns in their own activity design with the researchers as a support structure and also without the researchers after the latter leaves the school. We will work on encouraging teachers to use the common language in pedagogical patterns for describing their new but effective practices.
References


A Clustering Methodology of Web Log Data for Learning Management Systems

Stavros Valsamidis¹, Sotirios Kontogiannis¹, Ioannis Kazanidis², Theodosios Theodosiou² and Alexandros Karakos¹

¹Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi, Greece // ²Accounting Department, Kavala Institute of Technology, Agios Loukas, 65404, Kavala, Greece // svalsam@ee.duth.gr // skontog@ee.duth.gr // kazanidis@teikav.edu.gr // theodosios.theodosiou@gmail.com // karakos@ee.duth.gr

(Submitted July 2, 2010; Revised January 14, 2011; Accepted April 1, 2011)

ABSTRACT

Learning Management Systems (LMS) collect large amounts of data. Data mining techniques can be applied to analyse their web data log files. The instructors may use this data for assessing and measuring their courses. In this respect, we have proposed a methodology for analysing LMS courses and students’ activity. This methodology uses a Markov CLustering (MCL) algorithm for clustering the students’ activity and a SimpleKMeans algorithm for clustering the courses. Additionally we provide a visualisation of the results using scatter plots and 3D graphs. We propose specific metrics for the assessment of the courses based on the course usage. These metrics applied to data originated from the LMS log files of the Information Management Department of the TEI of Kavala. The results show that these metrics, if combined properly, can quantify quality characteristics of the courses. Furthermore, the application of the MCL algorithm to students’ activities provides useful insights to their usage of the LMS platform.

Keywords

E-learning, Web mining, Clustering, Metrics

Introduction

Learning Management Systems (LMSs) offer a lot of methods for the distribution of information and for the communication between the participants on a course. They allow instructors to deliver assignments to the students, produce and publish educational material, prepare assessments and tests, tutor distant classes and activate archive storage, news feeds and students’ interaction with multimedia. They also enhance collaborative learning with discussion forums, chats and wikis (Romero et al., 2008a).

Some of the most well-known commercial LMS are Blackboard, Virtual-U, WebCT and TopClass, while Moodle, Ilias, Claroline and aTutor are open source, freely distributed LMSs (Romero et al., 2008a). In Greece, the Greek University Network (GUNet) uses the platform Open eClass (GUNet, 2009), which is an evolution of Claroline (Claroline, 2009). This system is an asynchronous distance education platform which uses Apache as a web server, MySQL as its database server and has been implemented in PHP. Open eClass is open source software under General Public Licence (GPL).

Due to the volume of data, one of the main problems of any LMS is the lack of exploitation of the acquired information. Most of the times, these systems produce reports with statistical data, which, however, don’t help instructors to draw useful conclusions either about the course or about the students; they are useful only for administrative purposes of each platform. Moreover, the existing e-learning platforms do not offer concrete tools for the assessment of user actions and course educational content.

Data and web mining

Data mining is the search for relationships and patterns that exist in large databases, but are 'hidden' among the vast amounts of data. It is part of the whole Knowledge Data Discovery (KDD) process. KDD is the complete set of processes for knowledge discovery in databases that aims at the detection of valid information and pattern recognition in raw data (Kantardzic, 2003). The classical KDD process includes 5 phases: data pre-processing, data
transformation, data mining, data visualization and data interpretation. The first two phases select and "clean" a given dataset. The next phase, data mining, is essential in the whole KDD process; through it non-trivial patterns in data are found with the use of algorithms. Data mining consists of such tasks as classification, clustering, time series discovery or prediction and association rules mining (Witten and Eibe, 2000).

Web mining (Srivastava et al., 2000) is a sub-category of data mining. Data mining techniques are applied to extract knowledge from web data. There are three main web mining categories from the used data viewpoint: Web content mining, Web structure mining and Web usage mining (Spiliopoulou, 1999; Kosala and Blockeel, 2000; Bing, 2007). Web content mining is the process used to discover useful information from text, image, audio or video data on the web. Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. Web Usage Mining (WUM) is the application that uses data mining to analyze and discover interesting patterns of user data on the web. The usage data records the user’s behavior when he/she browses or makes transactions on the web site. The first web analysis tools simply provided mechanisms to report user activity as recorded in the servers. Using such tools, it is possible to determine such information as the number of accesses to the server, the times or time intervals of visits, as well as the domain names and the URLs of users of the Web server. However, in general, these tools provide little or no analysis of data relationships between the accessed files and directories within the Web space.

Data mining in e-learning

Data mining techniques have been used to discover the sequences patterns of students’ web usage after the analysis of log files data (Romero et al, 2007). Server log files store information containing the page requests of each individual user. After the pre-processing, phase of data analysis, this information can be considered as a per-user ordered set of web page requests from which it is possible to infer user navigation sessions. The extraction of sequential patterns has been proven to be particularly useful and has been applied to many different educational tasks (Romero et al, 2008b).

In this work, a methodology is proposed for the creation of a software tool which will be incorporated in e-learning platforms. With it students’ usage analyses can be made in order to motivate instructors to increase the use of platform for the needs of their courses. Instructors can benefit from the evaluations resulting from the proposed methodology, when trying to get a good place in the ranking of course usage and to improve their courses according to the tool indications at the relative units of the content. This improvement in the educational content will allow students to profit from the asynchronous study of courses by using actualized and optimal educational material. Next we describe work related to our research. Then our methodology is presented in detail followed by the experiments and the results. Finally we discuss the proposed methodology and describe future work.

Related work

Data mining may be applied to (i) traditional educational systems, (ii) web based courses, (iii) LMSs like Moodle, CLAROLINE, WebCT, etc., as well as (iv) adaptive and intelligent educational systems. Romero and Ventura (2007) surveyed several research studies in e-learning environments. One of their conclusions was that there is still not a standardized methodology for applying data mining techniques in this field. This lack of methodology gave us the motivation to propose a new approach with the different use of existing techniques.

Traditional educational data sets are normally small (Hamalainen et al, 2006) compared to files used in other data mining fields such as e-commerce applications; these may involve thousands of clients (Srinivasa, 2005). This is due to the typical, relatively small size of the classroom, although it varies depending on the type of the course (elementary, primary, adult, higher, tertiary, academic and special education); corresponding transactions are therefore also fewer. The user model is also different in both systems (Romero & Ventura, 2007).

There are several commercial tools such as DB Miner, Speed Tracer, Commerce Trends, Clementine, etc. (Galeas, 2009) and several free tools for WUM, like the Analog, Page Gather and SUGGEST.
The Analog system (Yan et al., 1996) consists of two main components, performing online and offline data processing with respect to the web server activity. Past users activity is recorded in server log files which are processed to form clusters of user sessions. The online component builds active user sessions which are then classified into one of the clusters found by the offline component.

Perkowitz and Etzioni (1999) proposed Page Gather, a WUM system that builds index pages containing links to similar pages. Page Gather creates index pages. The main hypothesis is that users behave coherently during their navigation. It deals with page clusters instead of session clusters, and bases them on the previous assumption called visit coherence, i.e. pages within the same session are in general conceptually related.

The SUGGEST WUM (Baraglia and Palmerini, 2002) system was designed to produce links to pages of the potential user interests. It can provide useful information to make web user navigation easier and to optimize web server performance. It was implemented as a module to Apache web server.

In addition to the above mentioned general purpose WUM tools, there are also several specialized ones that are used in the e-learning platforms. CourseVis (Mazza and Dimitrova, 2007) is a visualization tool that tracks web log data from an LMS. By transforming this data, it generates graphical representations that keep instructors well-informed about what precisely is happening in distance learning classes. Gismo (Mazza and Milani, 2004) is a tool similar to CourseVis, but provides different information to instructors, such as students’ details in using the course material. Sinergo/ColAT (Avouris et al., 2005) is a tool that acts as an interpreter of the students' activity in a LMS. Mostow et al. (2005) provides a tool which uses log files in order to represent the instructor-student interaction in hierarchical structure.

MATEP (Zorrilla and Álvarez, 2008) is another tool acting in two levels. First, it makes a mixture of data from different sources suitably processed and integrated. These data originate from e-learning platform log files, virtual courses, academic and demographic data. Second, it feeds them to a data webhouse which provides static and dynamic reports. Sinergo/ColAT (Avouris et al., 2005) is a tool that offers interpretative views of the activity developed by students in a group learning collaborative environment. Mostow (Mostow et al., 2005) describes a tool that shows a hierarchical representation of tutor-student interaction taken from log files.

An automatic personalization approach is also proposed by Khribi et al. (2009). It provides online automatic recommendations for active learners without requiring their explicit feedback through two modules: an off-line module which preprocesses data to build learner and content models, and an online module which uses these models on-the-fly to recognize the students’ needs and goals, and predict a recommendation list.

All these tools are based on the analysis of log files as our methodology does. Especially the Analog system (Yan et al., 1996) and the last one proposed by Khribi et al. (2009) seeded the idea for a final tool acting in two levels: online and off-line. However, none of the aforementioned tools proposes and uses indexes calculated by the pages and sessions accessed by the users. These indexes derive after the pre-processing of the raw data contained in the log files.

Methodology

The proposed methodology consists of three main steps, namely the logging step, the pre-processing and the clustering step. These steps are based on the framework described in detail in a study by Kazanidis et al. (2009) and facilitate the extraction of useful information from the data logged by a web server running an LMS. Instructors can benefit from the methodology’s course evaluation indexes.

The main advantages of the proposed methodology are that: (i) it uses data mining techniques for user and course evaluation; (ii) it proposes new indexes and metrics to be used with data mining algorithms; (iii) it can be easily adapted to any LMS, (iv) it visualizes the results in a user friendly environment and allows interactive exploitation of the data.
Logging the data

This step involves the logging of specific data from e-learning platforms. In more detail, the data recording module is embedded in the web server of the e-learning platform and records specific e-learning platform fields. Specifically, eleven (11) fields (request_time_event, remote_host, request_uri, remote_logname, remote_user, request_method, request_time, request_protocol, status, bytes_sent, referer, and agent) and user requests from different courses are recorded with the use of an Apache module, developed in Perl programming language, as a first step.

The Apache web page server uses the following configurations in its log files: Common Log Format, Extended Log Format, Cookie Log Format and Forensic Log Format. We used the latter configuration, because of the advantage that it stores server requests, before and after server process. Thus, there are two records per client request in a client's log file. For the recognition of each request, one unique ID is assigned to each request and a pair of signs (+/-), that signals the first or second record of a request.

The development of such a module has the following advantages: (i) rapid storage of user information, since it is executed straight from the server API and not by the e-learning application, and (ii) the produced data are independent of specific formulations used by the e-learning platform.

Data pre-processing

The data of the log file contain noise such as missing values, outliers etc. These values have to be pre-processed in order to prepare them for data mining analysis. Specifically, this step filters the recorded data delivered from step 1. It uses outlier detection and removes extreme values. This step is not performed by the e-learning platform and thus can be embedded into a variety of LMS systems.

The produced log file, from the previous step, is filtered, so it includes only the following fields: (i) courseID, which is the identification string of each course; (ii) sessionID, which is the identification string of each session; (iii) page Uniform Resource Locator (URL), which contains the requests of each page of the platform that the user visited. Although, these fields contain information about the e-learning process, more indexes and metrics (Table 1) are proposed in order to adequately facilitate the evaluation of course usage. However, although there are many metrics in web usage analysis for e-commerce (Lee et al., 1999), there is lack of corresponding metrics in e-learning. A simple approach was done by (Nagi et al., 2008) with “Reports”, which was embedded into the Moodle LMS for the courseware evaluation quality and the student interaction with the system. The ratio views/posts indicates the quality of the activity of a course, where "view" means that the data about access to an object is not saved into the database whilst "posts" means anything new that was created and uploaded, is saved in the database. Also, usability metrics for e-learning were proposed by Wong et al. (2003), where fuzzy systems were used to model each one of several factors and to reflect how each affects the overall value of e-learning usability.

Table 1. Indexes name and description

<table>
<thead>
<tr>
<th>Index name</th>
<th>Description of the index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sessions</td>
<td>The number of sessions per course viewed by users</td>
</tr>
<tr>
<td>Pages</td>
<td>The number of pages per course viewed by users</td>
</tr>
<tr>
<td>Unique pages</td>
<td>The number of unique pages per course viewed by users</td>
</tr>
<tr>
<td>Unique Pages per CourseID per Session (UPCS)</td>
<td>The number of unique pages per course viewed by users per session</td>
</tr>
<tr>
<td>Enrichment</td>
<td>The enrichment of courses</td>
</tr>
<tr>
<td>Disappointment</td>
<td>The disappointment of users when they view pages of the courses</td>
</tr>
<tr>
<td>Interest</td>
<td>It is the one ’s complement to the disappointment</td>
</tr>
<tr>
<td>Quality index</td>
<td>It represents the quality of the course combining Enrichment and Interest</td>
</tr>
<tr>
<td>Final score</td>
<td>It is the final score of the course</td>
</tr>
</tbody>
</table>
First, the number of the sessions and the number of the pages were counted in order to calculate course activity. The **unique pages** index measures the total number of unique pages per course viewed by all users. Unique pages are a set of pages uniquely identified by user sessions per LMS course. The **Unique Pages per Course per Session (UPCS)** index expresses the unique user visits per course and per session; it is used in order to calculate activity in an objective manner. For example, some novice users may navigate in a course visiting a page more than once. UPCS eliminates duplicate page visits, considering the visits of the same user in a session only once. UPCS was first introduced in the framework (Kazanidis et al., 2009) with a slightly different name, **UniquePCSession**.

**Enrichment** is a new metric which is proposed in order to express the “enrichment” of each course in educational material. We defined as the LMS enrichment index the complement of the division of unique pages per LMS course to total number of course web pages.

\[
\text{Enrichment} = 1 - \frac{\text{Unique Pages}}{\text{Total Pages}}
\]  

(1)

where Unique Pages $\leq$ Total Pages.

Enrichment values are between [0, 1). In a course with a minimal number of unique pages it is close to 1, and in a course where students follow only unique server paths it is 0. The enrichment metric shows how much information included in each course is handed over to the end user. It offers a measure of how many unique pages were viewed by the users in a course rich in educational material. If the enrichment of a course is less than the average of all courses, then the existing pages of the course should be enriched with more educational material. The enrichment metric was also first presented in the framework (Kazanidis et al., 2009) and was defined as the division of total course pages over unique course pages. However, in order to provide meaningful results in the range [0, 1), we redefined as proposed by Valsamidis et al. (2010a; 2010b).

We propose another metric called **disappointment**, which combines sessions and pages viewed by users. It measures how often when a user views a few pages of a course, s/he logs out the course. LMS disappointment metric is defined as the rate of sessions per LMS course to total number of course web pages.

\[
\text{Disappointment} = \frac{\text{Sessions}}{\text{Total Pages}}
\]  

(2)

A conceptual emotion detection and analysis system for e-learning using opinion mining techniques was presented by Binali et al. (2009). It has been discovered that disappointment can affect the e-learning experience. The disappointment metric reflects how quickly the users discontinue viewing pages of the courses. We define **interest** which is the complement to disappointment, because we want to use a positive sounding metric.

\[
\text{Interest} = 1 - \text{Disappointment}
\]  

(3)

Interest mining in virtual learning environments was presented by Gu et al. (2008). If differs from ours since it mines user interest mining via behaviour analysis. Both disappointment and interest metrics were proposed in Valsamidis et al. (2010a).

A low interest in a course means that there are not many unique pages viewed per session CourseID per Session; therefore the course is not quite popular among the students. This may be so either because students were not pleased with the educational material or there are not many pages to visit. In case that the Unique Pages was less than the computed average then the primary drawback is the low number of pages that the course contains. Otherwise, the problem has to do with the poor quality of the educational material.

The aforementioned metrics once counted allow us to rank courses. Specifically, **Quality index** is defined as the average of Enrichment and Interest. **Final score** is calculated as the product of Quality index with UPCS. It is worth mentioning that we tried to calculate Final score using different weights in Enrichment and Interest, but finally decided that it was fairer to use equal weights.

**Clustering**

In order to further explore the information provided by the LMS and the Apache log files and uncover patterns in the users’ activity, we applied a clustering method called **Markov Clustering algorithm (MCL)**. MCL separates users into different groups according to the usage patterns extracted from the e-learning platform. The MCL algorithm is used
to discover clusters in a graph. It performs an unsupervised clustering (no need to define number of clusters) based on weighted graphs and it has previously been used successfully in biology (Enright, 2002) and text mining (Theodosiou et al., 2008).

MCL is based on the idea that natural clusters in a graph have many edges between the members of each cluster and few across clusters. Once inside a cluster a hypothetical random walker will have little chance to get out of the cluster. MCL simulates random walks (flow) within the whole graph and strengthens flow where it is already strong and weakens it where it is weak. After many iterations of this process the underlying cluster structure of the graph gradually becomes visibly. Regions of the graph with high flows that describe clusters are separated by boundaries with no flow.

The MCL simulates the random walks within a graph by two algebraic operations called expansion and inflation that are applied to a stochastic matrix. The matrix representing the graph is used as input, and expansion and inflation are applied for many rounds until little or no changes are made in the matrix. The final matrix then represents the clustering of the graph nodes. Expansion refers to the power of a stochastic matrix using the normal matrix product. Inflation is the entry-wise Hadamard-Schur product (Radhakrishna & Bhaskara, 1998) combined with diagonal scaling and is responsible for both the strengthening and the weakening of the flow. The value of inflation controls the granularity of the clusters. The MCL algorithm is considered to be very fast and scalable, since its worst case time complexity is $O(N^3L^2)$, where $N$ is the number of documents and $L$ is an MCL parameter usually between 500 and 1000. The space complexity is $O(N^3L)$ (Dongen, 2000).

An example is presented in Figure 1. The original dense graph (upper left corner) is transformed in each step of the algorithm by pruning weak connections and keeping strong ones. Finally the algorithm convergences to a state where no more connections are pruned and the remaining ones formulate each cluster (lower right corner of the Figure 1).

Since MCL is applied to a graph we also draw a graph from our data. Each node of the graph represents a user. Two users are connected if they have both visited at least one common web page of the e-learning environment. The weight of the connection can be the number of common web pages.

![Figure 1. An illustrated example of the MCL algorithm](image)

**Visualization**

The final step of the methodology involves the visualisation of the clustering. This is achieved with *BioLayout Express 3D* (Goldovsky et al., 2005) that also allows the user to interact with results, for instance by searching for keywords, highlighting relevant documents, analyzing graph connectivity, linking nodes with external databases, and so forth. *BioLayout* is also open source, free software implemented in Java, that can be easily run under different operating system and thus it is ideal for our methodology.
BioLayout uses a modified version of the Fuchterman and Rheingold graph layout algorithm. It is used to produce an “aesthetically pleasing” layout of complex graphs (Enright et al., 2002). Based on this algorithm two users that are similar, meaning they have a connection with a high number of web pages that they both have visited, will end up closer in the final graph than two users which are weakly similar. Highly connected groups of similar users will form a tight cluster in the final graph.

It must also be noted that the MCL algorithm can be run under the BioLayout environment directly using the graph currently loaded in BioLayout and the resulting clusters are automatically coloured to help distinguish them.

In order to annotate each cluster, in a clutter-free way, we use the UserID and the CourseIDs he/she has visited to label each node of the cluster. This annotation schema allows the end-user to better understand the information in each cluster. Using the search capabilities of BioLayout the user can also easily find specific Courses in the graph and the users and groups related to them.

**Experimental scenario**

In order to evaluate the proposed methodology and assess its usefulness as a data mining tool for LMS systems, we used a dataset from a real e-learning environment.

**Dataset description**

The dataset was collected from the Technological Education institute (TEI) of Kavala that uses the Open eClass e-learning platform (GUNet, 2009). The data are from the spring semester of 2009 from the Department of Information management and involve 1199 students and 39 different courses. The data are in ASCII form and are obtained from the Apache server log file.

**Preprocessing**

The log file produced in the previous step is filtered, so it includes only the fields: (i) courseID, (ii) sessionID, and (iii) page Uniform Resource Locator (URL), as described in detail in the methodology section. Table 2 presents the results for 20 courses. The first 10 courses are with the highest Final Score, the next 5 courses are the ones ranked in the middle and the last 5 are the lowest ranked. We wanted to test our metrics, in best, average and worst cases from a usage point of view.

<table>
<thead>
<tr>
<th>Course ID</th>
<th>Sessions</th>
<th>Pages</th>
<th>Unique pages</th>
<th>UPCS</th>
<th>Interest</th>
<th>Enrichment</th>
<th>Quality index</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMD105</td>
<td>91</td>
<td>297</td>
<td>11</td>
<td>216</td>
<td>0,694</td>
<td>0,963</td>
<td>0,828</td>
<td>178,91</td>
</tr>
<tr>
<td>IMD35</td>
<td>87</td>
<td>338</td>
<td>8</td>
<td>179</td>
<td>0,743</td>
<td>0,976</td>
<td>0,859</td>
<td>153,84</td>
</tr>
<tr>
<td>IMD132</td>
<td>152</td>
<td>230</td>
<td>7</td>
<td>184</td>
<td>0,339</td>
<td>0,970</td>
<td>0,654</td>
<td>120,40</td>
</tr>
<tr>
<td>IMD36</td>
<td>72</td>
<td>217</td>
<td>7</td>
<td>134</td>
<td>0,668</td>
<td>0,968</td>
<td>0,818</td>
<td>109,61</td>
</tr>
<tr>
<td>IMD129</td>
<td>75</td>
<td>209</td>
<td>6</td>
<td>131</td>
<td>0,641</td>
<td>0,971</td>
<td>0,806</td>
<td>105,61</td>
</tr>
<tr>
<td>IMD125</td>
<td>93</td>
<td>164</td>
<td>6</td>
<td>134</td>
<td>0,433</td>
<td>0,963</td>
<td>0,698</td>
<td>93,55</td>
</tr>
<tr>
<td>IMD41</td>
<td>98</td>
<td>185</td>
<td>8</td>
<td>129</td>
<td>0,470</td>
<td>0,957</td>
<td>0,714</td>
<td>92,04</td>
</tr>
<tr>
<td>IMD66</td>
<td>56</td>
<td>144</td>
<td>9</td>
<td>107</td>
<td>0,611</td>
<td>0,938</td>
<td>0,774</td>
<td>82,85</td>
</tr>
<tr>
<td>IMD17</td>
<td>53</td>
<td>206</td>
<td>11</td>
<td>89</td>
<td>0,743</td>
<td>0,947</td>
<td>0,845</td>
<td>75,17</td>
</tr>
<tr>
<td>IMD111</td>
<td>33</td>
<td>142</td>
<td>9</td>
<td>79</td>
<td>0,768</td>
<td>0,937</td>
<td>0,852</td>
<td>67,32</td>
</tr>
<tr>
<td>IMD115</td>
<td>10</td>
<td>73</td>
<td>12</td>
<td>42</td>
<td>0,863</td>
<td>0,836</td>
<td>0,849</td>
<td>35,67</td>
</tr>
<tr>
<td>IMD9</td>
<td>26</td>
<td>105</td>
<td>12</td>
<td>42</td>
<td>0,752</td>
<td>0,886</td>
<td>0,819</td>
<td>34,40</td>
</tr>
<tr>
<td>IMD120</td>
<td>38</td>
<td>80</td>
<td>3</td>
<td>46</td>
<td>0,525</td>
<td>0,963</td>
<td>0,744</td>
<td>34,21</td>
</tr>
<tr>
<td>IMD112</td>
<td>30</td>
<td>62</td>
<td>6</td>
<td>46</td>
<td>0,516</td>
<td>0,903</td>
<td>0,710</td>
<td>32,65</td>
</tr>
</tbody>
</table>
We considered the evaluation of the courses primarily by the UPCS which is a quantitative metric. It is a quantitative metric because it just counts the number of instances. Courses with a high number of UPCS are quite popular among the students.

Since there are courses with the same UPCS (or with values very close to each other), we wanted to refine the situation and add an absolutely qualitative metric, named Quality index, which combines properly Enrichment and Interest. The final score derives as the product of UPCS with the Quality index.

**Clustering results**

The next step according to the proposed methodology involves the clustering of the data. In this phase we performed two distinct clusterings. The first one clusters the courses based on the proposed metrics described in the previous step. The second one clusters the students according to the courses they have in common. Thus, the proposed methodology provides the instructor with insight not only into the courses, but also into the students.

**Course clustering**

The clusters of the courses allow us to better assess the information contained in the proposed metrics. The clustering was performed using the open source data mining tool Weka (Weka, 2009). The metrics and indices described in the previous step were used with the SimpleKmeans for clustering platform courses. The properties of SimpleKmeans were Euclidean distance with 2 clusters, since our goal was to separate the 39 courses into high activity and low activity ones. The produced results show that 9 (23%) of the courses had high activity and 30 (77%) of them low activity.

Visualization of the results using UPCS is shown in Figure 2. Grey points in the left indicate high activity courses whereas points in black show low activity courses. It is evident that the information contained in the aforementioned metrics contains important information about the courses and can distinguish them in distinct groups.

![Figure 2. Cluster visualization using UPCS](image-url)
As shown in figure 3, all metrics converge as the results shown in figure 2 in two clusters. The high activity courses (grey) have high and medium final score, whereas the low activity courses have a low final score (black). Thus, the final score which combines properly all the proposed metrics, suffices to rank and eventually cluster the courses.

**Figure 3. Cluster visualization using Final Score**

**Student clustering**

Students are clustered using the stochastic algorithm called MCL as mentioned in the methodology section. The MCL is a graphed clustering algorithm and thus the relationships between the students are represented as a graph. The nodes of the graph represent each student and are labelled with its UserID and the CourseIDs s/he has visited. The vectors (unweighted) of the graph represent a relationship between two or more students. In order for the students to be connected, they have to visit at least one common course. The graph is represented in 3D using BioLayout as shown in Figure 4 (label nodes not shown for aesthetical reasons).

**Figure 4. The 1199 students and their relationships**

It is evident from the graph of Figure 4 that most of the students are highly interconnected (clustering of nodes that are represented as balls). Nevertheless, there is a small group of four students corresponding to course “Expert Systems” (CourseID=87) which is being taught at the last semester and only a very small number of students usually select and attend it. The students of this course are distinguished in the graph (Figure 4) even before applying the MCL algorithm, since they do not share another common course with the rest of the students. Thus, the graph
representation of the relationships of the students before the clustering can depict isolated students and facilitates the teacher to pinpoint specific groups of them.

BioLayout allows the instructor to interact with the graph and explore it, using a user friendly GUI. The end user can search for specific keywords, i.e. UserIDs or CourseIDs, and highlight relevant nodes. The user can also apply the MCL algorithm from the BioLayout environment. After the clustering is completed, BioLayout colours the nodes of each class differently in order to visually distinguish each cluster (Figure 5).

![Figure 5. The clusters of students in BioLayout using the MCL algorithm with inflation value 2.6](image)

After exploring the 3D graph with the help of BioLayout we applied the MCL algorithm with inflation value 2.6 to the 1199 students we collected from the Apache log file. MCL produced 27 clusters (Figure 5). We can see again that the four-student group for course “Expert Systems” are kept together forming one solid cluster. A detailed view of this cluster (Cluster 18) can be seen at Figure 6. We can see the UserID for each student, the CourseID and the number of incoming and outgoing connections for each node (incoming and outgoing connections link together users that visited the same pages), although in our case only the number of connections of the node in total (incoming and outgoing) is significant, since our graph is not directed. The rest of the students that were highly interconnected are split to the rest of the 26 clusters.

![Figure 6. The four student cluster of course “Expert Systems”](image)

By way of example, Figure 7 shows the detailed results for cluster 8. Cluster8 contains students that mainly attend the course “Mathematics in Management” (CourseID=60) and the course “Discrete Mathematics – Linear Algebra” (CourseID=64) which are taught at the first and second semester respectively. This can be explained since both courses are about mathematics, which have several common or similar exercises. Furthermore, the syllabus of each course is quite similar and the course “Discrete Mathematics – Linear Algebra” specializes in specific topics of the course “Mathematics in Management” that the students have been taught in the first semester.
Another example can be seen at Figure 8 where the results for Cluster 9 are presented. The students of this cluster have in common mainly the theory (CourseID=48) of course “Analysis and Design of Algorithms” and its related lab course (CourseID=49) that are taught in the third semester. The MCL managed to isolate these students from the total of 1199 students.

Discussion and conclusion

We propose a methodology to improve the content quality of LMS courses with the application of new metrics and existing clustering algorithms such as SimpleKmeans and MCL. We show that the proposed metrics can offer a preliminary course ranking, which in turn can be used as input for clustering algorithms. These suggest specific actions to instructors so that they can improve their course content, course usability and help them adapt their courses to student capabilities. Furthermore, we show that the MCL algorithm can successfully group together students who
visit courses with related content, based on student course paths in the LMS. In the future we will perform aggregate rankings on these groups, which are based on the proposed metrics (Enrichment / Interest / UPCS).

Specifically our methodology has the following advantages:
1. It is independent of the LMS platform since it is based on the Apache log files and not the LMS platform itself. Thus, it can be easily implemented for every LMS.
2. It uses new metrics in order to facilitate the evaluation of each course in the LMS and allows the instructors to make proper adjustments to their course e-learning material.
3. It uses clustering techniques in order to identify different groups of courses and different groups of students.
4. The MCL graph clustering algorithm is fast and scalable, and is applied for the first time in the field of e-learning. MCL can stack successfully courses into the same cluster based on the students’ course visits and can distinguish isolated courses.
5. The BioLayout visualization tool helps depict the clustering results and allows the instructor to further explore the results in an interactive 3D environment.

We received feedback about the methodology by the educators. The educators were informed about the indexing results and most of them increased the quality and the quantity of their educational material. They improved the quality by reorganizing the educational material in a uniform, hierarchical and structured way. They also increased the quantity by embedding additional educational material. An important outcome through the process of informing the educators about our results was that the ranking of the courses constitutes an important motivation for the educators to try to improve their educational material. Because of their mutual competition, they each want their courses to be highly ranked. However, a few educators complained that the organization of their courses does not assist them in having high final scores in the ranking list. They argued that, for example, the metric interest is heavily influenced by the number of web pages used to organize the educational material. Thus, courses that have all their educational material organised in a few pages have a low interest score. They were asked again to re-organize the material for each course in the LMS according to the order they are taught, in order to facilitate the use by the students. In the future we plan to evaluate our methodology by using a pre- and post-test design. A statistical comparison of the metrics of the courses before and after an instructor's exposure to and utilization of the metrics would offer valuable insights and help systematically evaluate our methodology.

We also plan to further automate the whole procedure, that is, we are developing a plug-in tool to automate the data pre-processing and clustering steps. This tool will run periodically (each month) and will e-mail to the instructors course ranking and suggestions. A similar policy was also applied by Feng et al. (2005), where in the long term, instructors were informed automatically by email about the quality of the content of their courses. We intend the final tool to offer insights at 2 levels: (i) On-line, with total statistical information such as number of visits per course (pages and sessions), student trends and activities at their visits, as well as detailed information per student (student duration per course and activity, student preferences and activities for all courses), and (ii) Off-line, with the use of data mining techniques such as pre-process, visualization, clustering, classification, regression and association, and discovering hidden data patterns.

References


**Blogging for Information Management, Learning, and Social Support during Internship**

Samuel K. W. Chu, Alvin C. M. Kwan and Peter Warning

Faculty of Education, Division of Information and Technology Studies, The University of Hong Kong, Hong Kong // samchu@hku.hk // cmkwan@hku.hk // pwarning@hku.hk

(Submitted August 9, 2010; Revised February 28, 2011; Accepted July 15, 2011)

**ABSTRACT**

The functions and possibilities afforded by blogging have been suggested to be relevant to learning and information management. Its increasing use in the business and education sectors is documented, but currently its use in professional education or internship is limited. The social nature of blogging appears to support the applicability of blogs to facilitate learning and communication between student interns. To contribute to the empirical evidence supporting these propositions, this study investigated the use of web logs to facilitate information management, learning, and mutual support for internship students. Undergraduate information management students ($N=53$) formed three cohorts who used three different blogging platforms in the course of their internships. They evaluated the use of blogs through an interview questionnaire that included close-ended rating scales and open-ended probes. The results revealed that students generally have positive perceptions on blogging as a tool to facilitate information management and mutual support. Blogs were also perceived to be useful in terms of self-reflection and communication. This study also showed that the blogging platform does not seem to have an impact on the students’ perceptions on the use of the blogs, whereas the communication between students and supervisors through the blog comments needs to be improved.

**Keywords**

Blogs, Internship, Information management, Reflection, Communication

**Introduction**

Blogs have existed since the 1990s, but their application to educational settings is a relatively recent phenomenon (Tan, 2010). A blog (abbreviation of “web log”) is a journal presented in reverse chronological order that consists of a person’s thoughts and ideas posted on the Web for multiple viewers (Flatley, 2005). It is a database containing text entries and other forms of content, including pictures and sound files which can be created, edited and published to the Web in a flexible manner. It has been suggested that blogging has attributes that are relevant to knowledge management [KM] (Levy, 2009). For instance, blogging has been identified as another means of storytelling, which is an established KM method.

Blogs have been effectively administered in higher education, as well as in distance learning settings (Buffington, 2007; Downes, 2004). Additionally, the application of blogs appears to have potential beyond classroom education. For instance, blogs may be a suitable tool for professional education or internship, since they are simple and accessible personal diaries that allow a community of users to interact and communicate with each other (Levy, 2009). The main purpose of an internship is to facilitate the application of theoretical concepts learned in classroom settings into professional practice, and in a manner, shifts an individual mind to a social setting (Gonczi, 2004). The social aspect of blogs seem to fit this perspective of internship; however, empirical evidence to support blogging for students during their professional education is still scarce.

**Blogging and knowledge sharing**

Knowledge is viewed as an output that emerges from active social construction (McDermott, 1999). The application of blogging for knowledge management has gained rapid recognition in business organizations, largely due to its potential as a channel for imparting knowledge, soliciting comments and links, and options for classifying and archiving entries (Ojala, 2005). Blogs have been found to be effective in organizing information, articulating, developing, and sharing ideas (Mortensen & Walker, 2002). Blogs can also act as tools for developing community relationships and conversational learning (Fiedler, 2003). Essentially, individuals can establish and maintain a personal network, and organize the information exchanged through blogging.
Knowledge sharing refers to the process by which knowledge from one individual is transformed, understood, and used by others (Ipe, 2003). While knowledge sharing has been explored in detail at the level of business organizations and task orientations, other forms of knowledge sharing occur over the Internet, such as blogs (Hsu & Lin, 2008). While effective information and knowledge sharing can be facilitated by a diverse range of Internet technology such as email and asynchronous discussion forums, blogs provide an environment that is relatively more advanced (Ferdig & Trammell, 2004). In contrast to a discussion forum that is shared by many, a blog provides full ownership to the user over his/her online content which may not fit neatly within a hierarchical or topic-based discussion forum.

**Blogging in education**

Interest in blogging as an educational resource has been fueled by its capacity for individual self-expression and social connectivity (Burgess, 2006). Blogs have been suggested to be more efficient than traditional pedagogical tools as they are portable, have automatic time and date stamps, can avoid misinformation due to time lag, and are economical if one has access to a computer and Internet connection (Gleaves, Walker, & Grey, 2008). The use of blogs as pedagogical tools has been documented with parents of first grade students (Clyde, 2005), with high school students (Downes, 2004), among librarians in higher education settings (Buffington, 2007), and in foreign language classrooms (Ducate & Lomicka, 2005).

Characteristics of blogging appear to support its use in the constructivist learning paradigm. Constructivist learning facilitates individuals to actively interpret and process information (Ally, 2008) in an emergent manner through their own inferences and discoveries (Villalba & Romiszowski, 2001). Social constructivist principles underpin the use of collaborative interactions among students to promote deep learning through exposure to alternative perspectives (Brett & Nagra, 2005). Two or more people learn something together through communication, negotiation, and production of materials (Gros, 2001). Recent studies indicate that collaborative learning is successful in online learning environments (Chu & Kennedy, 2011; Curtis & Lawson, 2001; Woo, Chu, Ho, & Li, 2011), and allows students to widen their knowledge base through interactions with other learners (Chu, 2008; Tyran & Shepherd, 2001). While collaborative learning emphasizes social and intellectual engagement (Smith & MacGregor, 1992), blogs have also been put forward as tools of social interaction and subsequent exploration among students through the mechanisms of comments and critiques.

Ellison and Wu (2008) have also proposed that blogging encourages critical and analytical thinking as it allows students to develop wider perspectives by interacting with their peers and even with a wider audience on the World Wide Web. Some findings revealed that peer feedback not only encourages learning among students but also leads them to engage in a higher level of understanding (Ertmer et al., 2007). Such a peer feedback system may be described as being inherent in blogging: some studies have found that the most effective aspects of maintaining a blog for students are reading others’ blogs, and reading feedback on one’s own blog (Ellison & Wu, 2008). The ensuing feedback mechanism may provide different perspectives from peers and encourage individual reflection (Lin, Hmelo, Kinzer, & Secules, 1999).

Reflection, which is an integral component of professional learning (Dietz, 1998), allows a learner to examine responses, beliefs, and context of experiences to gain new understanding (Rogers, 2001). It has been suggested that written records facilitate better recall and reflection than face-to-face discussions (Wells, 1999) such as the case of reflective journals. Blogs have been promoted as online forms of reflective journals (Richardson, 2005) and a number of studies have found blogs to be efficient reflective tools (Lin et al., 1999; Stiler & Philleo, 2003). Furthermore, blogs have also been shown to be good motivational tools that encourage students to engage in interactive discussions (Chu, Chan, & Tiwari, 2012; Downes, 2004; Flatley, 2005).

Internship has long been recognized as an important method to prepare students for professional careers. Hands-on experience, informal apprenticeships with active professionals, and exploration of learning through other institutions (e.g., business, community associations) are fundamental ways of refining students’ skills and helping them attain new insights on their professions (Weinberg, 1986). Since internship is an effective transitory phase towards professional development, a significant body of literature has focused on the theories of design and experience in its implementation (Lloyd & Bristol, 2006). Murray-Harvey (2001) investigated sources of student support from teachers during a training program; the findings revealed that students placed a high value on emotional support from
teachers, as well as collaboration and feedback from the teachers. In many cases, it is difficult to provide support on a face-to-face basis due to time constraints and geographical dispersion. Some studies have explored the use of information technology for providing support during students’ internship. These studies examined web-based asynchronous discussions (Doering, Hughes, & Huffman, 2003), email communications (Graf & Stebnicki, 2002), and the use of video-conferencing (Nasiopoulos & Ward, 2002).

It is apparent that the potential of blogging as a resource for information and knowledge sharing has been examined substantially in both business organizations and conventional education settings. However, little is known about the use of blogs as a supportive tool during internship. Using the framework of collaborative learning, this study utilized blogging as a platform for reflections during the internship of undergraduate students in the B.Sc. in Information Management (BScIM) programme of the University of Hong Kong. Building on the growing evidence that supports the suitability of blogging for constructivist learning approaches, we explored the potential of blogging during internship with foci on information and knowledge sharing, as well as social support. The blogging behaviors and perceptions of three cohorts of students were examined using quantitative and qualitative methods.

**Research methods**

This study aimed to describe the students’ blogging behaviors, and determine their perceptions of blogging in terms of facilitating information management, mutual support, learning, and communication during internship. All the procedures were reviewed and approved by the ethics review committee of the university.

**Participants**

Undergraduate students \(N=53\) in the BScIM programme were grouped in three cohorts based on their years of enrolment from 2006 to 2008. They all took part in an internship programme lasting for one to three months in various organizations. Research participants included 16 out of 19 interns in 2006, 16 out of 21 interns in 2007, and all the 21 interns in 2008.

**Procedures**

Undergraduate students in the BScIM programme undergo internship in their third (final) year over a period of one to three months. They are assigned to different organizations of varied nature (e.g., library, hospital, police force, publishers, and law firms). For each year, approximately 80% of the students undertook their internship in local organizations in Hong Kong, while the remaining 20% did their internship in Mainland China or in overseas organizations. The students’ tasks revolve around information management, but may vary in terms of scope and level of technical requirements.

At the onset of their internship, students were asked to create blogs as a means of sharing their experiences and posting reflections. Internship requirements included the submission of self-reflections, and the students were asked to write their self-reflections and post them as blogs every one to two days, for the duration of their placement. Students were advised to share what they have learnt, and useful resources they have found, with their classmates. Each cohort of students had a different blogging environment. The first cohort used a commercial blogging system: either Xanga (http://hk.xanga.com) or Blogger (https://www.blogger.com). The second cohort used an open source content management system named Drupal, while the third cohort used another commercial blogging system called YouBlog (http://www.youblog.cc). The free commercial blogging systems had common functionalities that supported blog creation, editing, viewing, commenting, file attachment, and links creation. Additionally, Xanga and YouBlog provided view counts for each blog entry. On the other hand, Drupal is an open-source, self-hosted content management system which includes blogging among its supported features. Both Drupal and Xanga allow users to provide feedback to bloggers by giving a rating for each blog entry.

After their internship was completed, students were invited to participate in telephone interviews where they reported their blogging frequency and rated the effectiveness of blogs for learning and support through the use of a 4-point Likert-type scale. The Likert-type scale utilized a score of 1 for “Strongly disagree”, 2 for “Disagree”, 3 for “Agree”, 4 for “Agree”, and 5 for “Strongly agree”.
and 4 for “Strongly agree”. Reasons to support their response choices were probed through open-ended questions. All the interviews were audio-recorded with the participants’ consent and transcribed for data processing.

**Data analysis**

Numerical data obtained from the questionnaires were analyzed using non-parametric tests (SPSS version 16.0). Responses on the Likert scales were summarized using descriptive statistics, and median scores that were higher than 2.5 were interpreted as edging towards the positive, while median scores that were lower than 2.5 were edging towards the negative feedback. Interquartile range was also calculated for the blogging frequency, and scores below the 25th percentile were classified as infrequent users, while scores above the 75th percentile were classified as frequent users. Statistical significance level was set at $p<0.05$. Qualitative data obtained from the open-ended questions were analyzed using NVivo version 7.0.

**Findings and discussion**

**Blogging behaviours**

A total of 871 blogs entries were identified in the data captured by the blogging systems, where the average number of blog entries for each student was 45.8 ($SD=28.5$). Users’ blogging frequency was reported in terms of the frequency of their blog-writing, reading their own and others’ blogs, and giving comments, as shown in Table 1.

<table>
<thead>
<tr>
<th>Blogging Behaviours</th>
<th>Cohort 1 (n=16)</th>
<th>Cohort 2 (n=16)</th>
<th>Cohort 3 (n=21)</th>
<th>Sig. Kruskal-Wallis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing one’s own blog</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>.484</td>
</tr>
<tr>
<td>Reading one’s own blog</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>.686</td>
</tr>
<tr>
<td>Reading classmates’ blogs</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>.195</td>
</tr>
<tr>
<td>Commenting on classmates’ blogs</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>.002*</td>
</tr>
<tr>
<td>Overall blogging behaviours</td>
<td>2.5</td>
<td>2.5</td>
<td>3</td>
<td>.421</td>
</tr>
</tbody>
</table>

*Statistically significant at $p < 0.05$

A general measure of the blogging behaviours was represented by the median of the ratings for the four blogging behaviours, which was taken for each participant and recorded as the “overall blogging behaviour”. The students seemed to prefer writing their own blogs and reading their own and others’ blogs rather than making comments on other students’ blogs. The first and second cohorts, in particular, rarely gave comments to their classmates’ blogs. Students wrote their blogs in lieu of the required self-reflection during their internship, and this may partly explain the frequency of writing their own blogs. On the other hand, writing comments on their peers’ blogs is not a requirement in the course which may have resulted in less motivation to comment. The commenting function in blogs, along with the flexibility in time and space, has been shown to facilitate critical reflection in learning (Yang, 2009). However, our findings indicate that among undergraduate interns, this outcome may require explicit motivators in order to be maximized. A significant difference between the three cohorts was observed: the third cohort commented on their classmates’ blogs more frequently. There were no systematic differences in the learning conditions between the three cohorts, except for the blogging systems that were used. The third cohort used YouBlog, which may have been generally easier to use, motivating interns to engage in more frequent blogging activities.

**Perceptions of blogging as a tool for information management and mutual support**

The perceived effectiveness of blogging as a tool for learning was evaluated in terms of facilitation of knowledge sharing, information sharing and social support among interns. Generally, the students gave ratings that were on the positive side of the scale for all three dimensions. Table 2 shows that the students appeared to recognize the effectiveness of blogging in facilitating information management and mutual support. It has been suggested that blogs are good for supplementing face to face interactions among graduate students (Buffington, 2007). The findings
from the students’ ratings appeared to support this proposition. No significant differences were noted between the ratings of the frequent and infrequent users. Previous studies have shown that learners who use an online technology for learning more frequently tend to give lower ratings for its effectiveness (Chu, 2008). This has been explained by the higher chances of noting the technology’s limitations as a user becomes more familiar with the technology. In this current study, the frequency of blogging did not seem to influence the interns’ perceptions on the effectiveness of blogging for learning. In the preceding section, the number of blog entries per student (\(M=45.8\)) has been reported, indicating that a substantial amount of blogging has been done by the interns. This may explain the generally consistent perceptions of students, regardless of their blogging frequency.

### Table 2. Students’ ratings on the effectiveness of blogging for information management and mutual support

<table>
<thead>
<tr>
<th>Variables measuring the effectiveness of blogging</th>
<th>Overall ratings</th>
<th>Frequent users</th>
<th>Infrequent users</th>
<th>Sig. Mann-Whitney</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Blogging facilitated knowledge sharing</td>
<td>2.8 (0.62)</td>
<td>2.93 (0.48)</td>
<td>2.6 (0.84)</td>
<td>0.61</td>
</tr>
<tr>
<td>Blogging facilitated information sharing</td>
<td>3.1 (0.53)</td>
<td>3.29 (0.47)</td>
<td>3.0 (0.56)</td>
<td>0.26</td>
</tr>
<tr>
<td>Blogging facilitated mutual emotional support</td>
<td>2.8 (0.70)</td>
<td>3.0 (.056)</td>
<td>2.64 (0.50)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

Note. 1Using the 4-point Likert scale as discussed in the methodology section.

It has been suggested that by accessing blogs of others with similar interests, bloggers can obtain information that may serve as useful resources (Efimova, 2004). In this study, the main benefits of blogging for information sharing reported by most students include free mutual reading allowing them to know what the others are doing, obtaining information regarding other job fields, and easy access to research-related information. One example of students who share information on blogs is extracted below:

“The seminar will be held again on 23 August, if you are interested, go to the Census and Statistics Department’s website for more details: http://www.censtatd.gov.hk/statistical_literacy/course_and_seminars/scientific_sample_survey/index_tc.jsp”

Blogs provide a virtual personal space, while at the same time allow online communities to form (Glogoff, 2005). Bloggers are able to share knowledge and express their personal styles (Nardi, Schiano, Gumbrecht, & Swartz, 2004). It appears that blogging provided a venue for interns to share knowledge based on their personal experiences. Probing questions in the interviews revealed further the participants’ explanations of their perceptions of blogs’ usefulness for information and knowledge sharing (see Table 3). Blogging has been reported to be useful for interns in terms of sharing information and knowledge on their respective work assignments. Interns also appreciated the ability to upload multi-modal forms of information (e.g., photos, videos) on the blogging platform. The potential of learning from their peers’ problem-solving experiences was noted by only two of the participants. Problem-solving that is shared in a community has been described as a representation of knowledge exchange and serves as a means of knowledge building (Yang, 2009). The varied nature of the organizations to which the interns were assigned resulted in disparities in their tasks. This may have resulted in interns’ limited perceived usefulness of the problem-solving experiences of their peers.

### Table 3. Students’ opinions on blog as a suitable tool to facilitate information management and mutual support

<table>
<thead>
<tr>
<th>Positive Feedback</th>
<th>Negative Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides knowledge on what others are doing (13)</td>
<td>Most blog entries are unimportant and routine (6)</td>
</tr>
<tr>
<td>Provides free mutual reading (9)</td>
<td>Most entries are personal and are not useful comments (4)</td>
</tr>
<tr>
<td>Provides information on other work fields (8)</td>
<td>Students just write little in order to complete the assignment (4)</td>
</tr>
<tr>
<td>Enables sharing of information on research, photos, videos (8)</td>
<td>Face-to-face communication is necessary (2)</td>
</tr>
<tr>
<td>Allows sharing of experiences (5)</td>
<td></td>
</tr>
<tr>
<td>Provides access to information everywhere, anytime (4)</td>
<td></td>
</tr>
<tr>
<td>Serves as a source for problem-solving mechanisms (3)</td>
<td></td>
</tr>
<tr>
<td>Provides opportunities to learn from others’ experience and analysis (2)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Students were asked to comment on the effectiveness of information and knowledge sharing by blogging
The primary negative feedback was related to the superficiality and routine nature of most blog entries. The students’ blogs represented their self-reflections which made up an academic requirement for their internship. It appeared that some students may not have engaged in deep reflection through their blogging. An earlier study indicated that blogging encouraged ongoing and thoughtful reflection by making thinking visible and available for critique and comment by others (Luehmann, 2008). However, as a tool for reflection, blogs have been observed to be more descriptive than critical (Yang, 2009). The findings of this study appear to be consistent with this latter observation, and may be partially due to the required nature of the interns’ blogs. Some interns may have written entries that were superficial and routine, primarily to fulfill the academic requirement. Nevertheless, it may not be discounted that the requirement may have pushed some other students to exert extra effort into writing their blogs to improve their academic standing. In any case, these aspects need further examination that is beyond the scope of this report. Future inquiries may investigate students’ intrinsic motivation to blog as opposed to blogging to fulfill the required reflection. This may reveal further insights on optimizing the educational impact of blogging. At this point, the findings essentially imply that there is a need to increase the students’ awareness of blogs’ potential usefulness during an internship.

Blogs were also found to enhance social support among students in various ways. Student C1-S6 (Cohort 1-Student 6) reported that his classmates did not only write about what they had learnt, but also about their relationships with colleagues and supervisors so that insights were gained on dealing with interpersonal relationships in positive ways. Some students used blogs to vent their frustrations at work. For instance, student C1-S4 complained about not getting any interesting tasks at work, while other students complained about the heavy workload or huge pressure they had during their internships.

Perceptions of blogging as a tool for learning

As a tool for learning, blogging was evaluated in terms of usefulness for reflection and communication, and students reported positive perceptions on these aspects (see Table 4). It has been previously suggested that blogs may also serve as online journals (Richardson, 2005) and studies have found them to be efficient reflective tools (Stiler & Philleo, 2003). It appears that interns recognized this benefit of blogs as well. The effectiveness of blogs for self-reflection has been described to be due to the easy retrieval of previous information (Flatley, 2005), automatic archiving of information, and more effective monitoring of one’s progress (Lin, et al., 1999). Students can always retrieve their reflections from their blogs by typing the date of entry. More importantly, this reflection process can contribute to effective learning as it enables students to better understand what they have learnt and how these learnt skills can be applied in the future (Tan, 2006).

Table 4. Students’ ratings on the usefulness of blogs for communication and learning

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>Mean (SD)</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogging is useful for self-reflection</td>
<td>3.2 (0.65)</td>
<td>3</td>
</tr>
<tr>
<td>A blog is suitable for recording self-reflections</td>
<td>3.0 (0.65)</td>
<td>3</td>
</tr>
<tr>
<td>Blogging is useful for communication among classmates</td>
<td>3.0 (0.5)</td>
<td>3</td>
</tr>
<tr>
<td>Classmates’ comments on blogs are helpful</td>
<td>2.1 (0.99)</td>
<td>3</td>
</tr>
<tr>
<td>Supervisors’ comments on blogs are helpful</td>
<td>0.6 (1.09)</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. 1Using the 4-point Likert scale as discussed in the methodology section.

Table 5. Students’ opinions on using blogs to keep self-reflections

<table>
<thead>
<tr>
<th>Positive Comments</th>
<th>Negative Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serve as reminders for projects (14)</td>
<td>Privacy concerns (4)</td>
</tr>
<tr>
<td>Record and reflect things systematically (12)</td>
<td>Concerns about grades and authenticity of content (3)</td>
</tr>
<tr>
<td>Convenient (7)</td>
<td>The mandatory use of computers is found to be problematic (2)</td>
</tr>
<tr>
<td>Entries are easy to trace (5)</td>
<td>There are better alternatives, e.g., Facebook (2)</td>
</tr>
<tr>
<td>Readily usable for sharing (5)</td>
<td></td>
</tr>
<tr>
<td>Easy to access (4)</td>
<td></td>
</tr>
<tr>
<td>Suggestions from classmates and supervisors are received (3)</td>
<td></td>
</tr>
<tr>
<td>Allow expression of feelings (2)</td>
<td></td>
</tr>
<tr>
<td>Centralized platforms (1)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Students were asked to comment on the blog’s suitability to keep reflections.
Table 5 summarizes the students' opinions on using blogs for self-reflection. It was noted that blogs were useful reminders to prevent students from missing data for their projects. Additionally, the participants reported that blogs provided a systematic record of what they have done at work, and they were convenient to use. The major negative comment on the blogs' use for self-reflection is the concern over privacy. Student C1-S7 commented on this by saying that "the blog is a platform that anyone can access, which information can be easily traced by date of entry and type of content. Outsiders would then be able to see information that I want to keep private." It is worth noting that such concerns have been adequately addressed by the blogging platforms by allowing users to customize their privacy settings to restrict access to selected postings.

Compared to traditional means of communication, blogs have been considered desirable tools for students to make their thoughts, feelings and experiences accessible to a wider audience (Ellison & Wu, 2008). Student C1-S10 suggested that benefits of communicating through blogs included increasing the transparency of work. Because the BScIM course requires real-life sharing of experiences from different approaches, this student further stated that communication through blogs allowed interns working overseas to communicate with each other as it provided a viable alternative to face-to-face interactions. Student C2-S8 worked overseas and thought that blogs were useful for communication as phone calls were seldom made to classmates who were working in Hong Kong. It appears that blogs were found to be particularly useful for students who are separated by geographical distance. Another student (C1-S6) described the blog as a centralized platform for communication which effectively countered the difficulties associated with the physical distance at work.

From the probing questions, a number of students did not think that blogs were a very suitable communication tool since their classmates or supervisors preferred using emails or direct conversation to communicate. Student C1-S3 clarified this by saying that "blogs might be more useful for students working abroad." However, for students who worked in Hong Kong, face-to-face interaction with their classmates and supervisors was preferred.

Participants gave unfavourable ratings for the helpfulness of comments on their blogs, including those from their classmates (mean=2.1±0.99) and supervisors (mean=0.6±1.09) respectively. Probing inquiry revealed that students perceived the comments from their peers as shallow and not useful (14 out of 53 participants), while a few of them remarked that suggestions from one work area may not be applicable to another (5 out of 53). Negative views on the usefulness of their supervisors’ blog comments were reported, mainly because they were either infrequent or absent (41 out of 53). The majority (89%) of the students reported that they received blog comments from their supervisors once per month or less. Student C3-S11 gave a typical response of the interns regarding supervisors’ comments “He (my supervisor) has never given me any feedback or comments via the blog.” In another interview, Student C3-S3 noted that, “my supervisor had never visited my blog.” Students who received comments from their supervisors regarded the comments as helpful advice for problem solving and academic work (9 out of 53). As some studies suggest that students have to learn through self-exploration and directive learning with their supervisor’s feedback (Glogoff, 2005), it appears that there is a need to improve the regularity of supervisors’ feedback in order to make learning more effective. The existence of supervisors’ comments serves as guidance for students, and also as an indicator that they consider blogging to be an important communication tool. This may potentially encourage students to blog more frequently. Table 6 summarizes the information obtained through the open-ended questions.

| Table 6. Students’ opinions on the usefulness of classmates’ and supervisors’ comments |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Positive Comments              | Negative Comments              |                                 |
| Comments from classmates       |                                 | Shallow and useless (14)         |
| Allowed difficulties to be shared; solutions and advice were obtained (9) | Seldom (if not never) received (12) |
| Provided encouragement (9)     | Comments regarding other work fields were not applicable (5) |
| Provided information on how the others are doing and their work areas (6) |                                 |
| Provided a reflection of work situations and job scope (4) |                                 |
| Different perspectives and viewpoints were discussed which helped problem solving (2) |                                 |
| Promoted resolution of disagreements and sharing of ideas (2) |                                 |
| Stimulated thinking (1)        |                                 |
Comments from supervisors
- Provided professional and academic advices that helped problem solving (9)
- Indicators that supervisors kept track of students’ progress (2)

Note. Students were asked to comment on the usefulness of comments from other students and supervisors.

While the blogging platforms differed for each cohort, our results show that there were no significant differences in the ratings on the usefulness of the blogs as a platform for learning among the participants (all \( p > .05 \), see Table 7).

Table 7. Comparing the 3 cohorts’ ratings on the usefulness of blogs as a platform for learning

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>Cohort 1 (n=16) Median</th>
<th>Cohort 2 (n=16) Median</th>
<th>Cohort 3 (n=16) Median</th>
<th>Sig. (Kruskal-Wallis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogging is useful for self-reflection</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>.46</td>
</tr>
<tr>
<td>Blog is suitable for recording self-reflections</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>.11</td>
</tr>
<tr>
<td>Blogging is useful for communication among classmates</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>.99</td>
</tr>
<tr>
<td>Classmates’ comments on blogs are helpful</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>.76</td>
</tr>
<tr>
<td>Supervisors’ comments on blogs are helpful</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.17</td>
</tr>
</tbody>
</table>

Note. Using the 4-point Likert-type scale as discussed in the methodology section.

While we noted in the earlier section of the findings that the blogging platform may have influenced the frequency of comments, the platform did not appear to influence the students’ perceptions on blogging. As noted above, consistent negative feedback was also given for the usefulness of classmates’ and supervisors’ comments. These findings imply that while the type of blogging platform may not have a significant effect on blogging impact during internship, the quantity and quality of comments from classmates and supervisors have a more important effect.

Conclusion and implications

In the constructivist learning paradigm, an active exchange and interpretation of information facilitates knowledge construction in learning (Vygotsky, 1986). Along the lines of this theoretical framework, blogging has been suggested to address the discursive nature of knowledge construction (Ferdig & Trammell, 2004). Our findings indicate that across three cohorts of students, blogging has been reported to be supportive of constructivist learning by facilitating information and knowledge sharing. While such positive perception was consistent, the usefulness of blog contents has been inconsistent among participants. Furthermore, comments from classmates and supervisors have also been reported as inadequate and shallow. Collaborative learning places a great emphasis on the extent and quality of the exchanges that occur among students in a given environment (Dillenbourg & Schneider, 1995). Our findings confirm that blogging is perceived to support collaborative learning, but there is a need for scaffolding that has to be addressed. Scaffolding has been used as a metaphor referring to approaches wherein teachers provide students with the tools they need to learn (Jacobs, 2001). Originally used in teaching children, scaffolding has also taken various forms of adult support in terms of providing demonstrations and guidelines, and keeping learners’ attention focused (McDevitt & Ormrod, 2002). Our findings suggest that blogging provides a platform for knowledge construction and given adequate scaffolding may facilitate constructivist learning during internship. Scaffolding for blogging in higher education may take on the form of consistent guidelines and recurrent reminders of the focus and purposes of the task. This extends current theoretical explorations of using technology to facilitate constructivist learning, and takes the focus towards internship in particular.

The blogging interface that generates feedback from readers naturally positions the students for reflective opportunities (Ferdig & Trammell, 2004). Regardless of the platform used, the utilities of blogging systems have allowed interns to record their experiences and generate responses from classmates and supervisors. Consequently, participants across cohorts confirmed the usefulness of blogging for reflection. A framework for the educational affordances of blogging has been suggested by Deng and Yuen (2011), pointing out that reflection occurs as a blogger writes, reads, and exchanges comments. While our quantitative findings confirm that students perceived writing blogs to be useful for self-reflection, findings from the probing questions indicate perceived benefits from
reading blogs (i.e., reminders, easy sharing) as well. In contrast, comments from classmates and supervisors were perceived to be of limited worth. We suggest that these findings support the framework of blogging affordances, where reflection involves reading, writing, and exchanging comments. In fact, students explained the dissatisfaction with comments to be due to lack of both depth and frequency. We thus infer that reflection is best facilitated by blogging when the educational affordances are harnessed into the system. The technological characteristics of blogging systems appear to be adequate to support these educational affordances. Our findings imply that the inadequacy in employing blogging’s educational affordances might be associated with pedagogical factors that are intrinsic in each particular internship program.

In the context of internship, our participants were required to write their reflections as blogs. Consequently, their blogging behaviours were more focused on writing their own blogs, and giving rise to comments that were described to lack depth. On the other hand, supervisors’ participation in the blogging activities may be constrained by limited time resources, a desire to limit their impact on students’ independence of judgement during internships, or other reasons. These factors associated with the context of internship appear to have important effects on how interns used blogging for learning. Earlier studies have also shown that the pedagogical context of a course influences students’ blogging behaviours (Kerawalla, Minocha, Kirkup, & Conole, 2009). In this study, students tended to direct their attention towards blog writing as a consequence of the internship context. While the nature of internship is distinct from a classroom-based coursework, our findings imply the value of the pedagogical context of learning in utilizing the educational affordances of blogging. Nevertheless, we acknowledge that our findings may have limited generalizability to interns in other disciplines or fields of study. Future research work on using blogs for internship is warranted to build on our findings.

Further studies might also examine the use of blogging as a learning tool during internship using other outcome parameters. Interns reported generally positive perceptions on blogging for information and knowledge sharing, self-reflection, social support, and communication for those who are placed overseas. While this is an encouraging basis for future use of blogging in education, there is no causal mechanism that has been established. Furthermore, the lack of association between frequency of use and interns’ perceptions imply that students’ blogging behaviours (i.e., amount of blogging activity) may relate to students’ views of blogging for learning. For educators, this brings about the question of the actual utility of blogging as a tool for learning, which may be addressed by further research that may delve into learning processes and outcomes. In conclusion, we found that in the context of internship, blogging was found to be a generally useful tool that supports constructivist learning. Future applications of blogging in education are also suggested to account for the pedagogical context which may include scaffolding strategies.

References


Using Mobile Learning to Improve the Reflection: A Case Study of Traffic Violation

Yu-Feng Lan and Shin-Ming Huang
Department of Information Management, National Formosa University, YunLin, Taiwan, R.O.C. // yflan@nfu.edu.tw

(Submitted August 10, 2010; Revised November 23, 2010; Accepted July 1, 2011)

ABSTRACT
The purpose of this study was to integrate mobile communication technologies and a global positioning system (GPS) to construct an instant, convenient report of the mobile network service system named the Mobile Traffic Violation Reporting System (MTVRS), to improve learners’ traffic violation reflection level. Data were collected using a “usefulness of mobile learning systems” questionnaire and traffic violation reflection questions from a sample consisting of 32 learners in Taiwan. Learners voluntarily participated in a four-week program using mobile devices to transmit photographs of local traffic violations and exchange pictures and observations. The experimental results revealed by the questionnaire were satisfactory and encouraging in terms of the effectiveness of the supported mobile learning environment. Moreover, the responses to the questionnaire showed significant differences in regard to both gender and age. The study also found a positive improvement in reflection levels toward traffic violations among groups following the experiment.

Keywords
Mobile learning, Reflection, Global positioning system (GPS), Traffic violation, Gender

Introduction
Thanks to advances in wireless and mobile technologies, and through the use of mobile devices, it is possible to extend the learning environment far beyond classroom walls (Liu, 2007). Educators have begun to consider mobile learning (ML) with the prospect of extending the realm of computer-mediated education to learning situations outside of traditional educational settings (Chen et al., 2003; Uzunboylu et al., 2009). These technologies provide the potential for collaborative interaction and learning opportunities for geographically dispersed persons and groups (Huang et al., 2009). Moreover, both learners and instructors can have a strong social impact outside the classroom through using mobile devices.

Learning via mobile devices is widely accepted by the learner community. Learners are interested in using all available ML resources through mobile phones and PDAs to access information anytime and anywhere (López et al., 2009). The aim of ML is to provide an educational environment in which learners can learn without any limitation of time, place, or device, thereby realizing a more creative and learner-centered educational process (Joo & Kim, 2009). Rogers et al. (2010) showed handhelds can accompany the learner into the field and can at the same time scaffold exploratory activities, for instance by enabling learners to sense and record aspects of the local environment, while giving opportunities for taking electronic notes and at the same time drawing on various digital resources and representations.

Furthermore, according to the findings of prior research, the use of mobile devices to support learning activities offers several benefits (Huang et al., 2008; Chu et al., 2010; Huang et al., 2010). These include the ability to: (a) improve communication and collaborative interaction, (b) provide more learning opportunities for geographically dispersed persons, (c) encourage active learning, (d) enhance the learner’s feedback process, and (e) acquire content quickly.

Although several studies have reported the benefits, implementation of mobile learning is not totally without problem. For example, Vavoula et al. (2009) presented an evaluation of a Myartspace website, a service on mobile phones for inquiry-based learning allowing learners to gather information during a school field trip. Their results showed the system service was effective in enabling students to gather information in a museum and this provided a novel application for construction and reflection in the classroom. Though Myartspace allows the user to take a photo and then record an audio or text comment about the photo, those items were not well linked by the software. As a result, the inquiry contents could not be successfully integrated and used within the website. Additionally,
Uzunboylu et al. (2009) examined the integration of mobile technologies, data services, and multimedia messaging systems to develop environmental awareness. The results showed the environmental awareness of students increased significantly, and attitudes toward maintaining clean environments and preventing pollution improved. Most importantly, in their conclusion they pointed out ML will likely provide a suitable platform for use in a wide range of educational settings. Accordingly, to investigate the effect of different applications of ML in educational settings is emerging as important issue.

**Motivation and significance of the study**

Every year, traffic accidents result in huge loss of life, assets and social cost. As the vehicle population continues to grow within a limited space, traffic-jams and traffic violations have occurred and need to be solved. Specifically, because most traffic incidents result from negligence or violations by road users, enhancing traffic education thus becomes very important. Therefore, Iversen (2004) examined whether attitudes toward traffic safety issues are predictors of future risky behavior in traffic. The results of research show a high correlation between the dimensions of attitudes and behavior.

Traffic education can be defined as the process including all rules which the people must know to be protected in traffic (Yilmaz & Celik, 2008). Generally, learners have more learning opportunities in educational environments. However, traffic education has not received enough attention and has been conducted on an irregular basis in schools. Besides, people usually understand what situations may cause a traffic violation problem; however, they are apt to make mistake carelessly. From the reflection perspective, Dewey (1933) pointed out reflection could support learners a persistent and careful thinking action during a problem solving process. In this study, traffic violation reflection can be viewed as the improving process in which the learners must understand what situations may cause a traffic violation problem, according to the traffic laws, and further propose possible ideas or solutions to avoid such violation.

To the best of our knowledge, no studies have been conducted on a relevant teaching approach to enhance traffic violation reflection based on the integration of the learning theory and mobile technology. Therefore, a better understanding of how to use existing technologies and learning theory to improve traffic education with limited time and resources, providing instructional resources/materials, and creating better traffic violation reflection, is worth examining.

**Theoretical foundations**

**Social constructivist theory**

Social constructivist theory assumes learners act and reflect within an environment and specifically emphasizes the importance of culture and context in understanding what occurs in society and constructing knowledge based on learners’ understandings (Derry, 1999). The learning mode is constructivist in approach, whereby the learners actively participate in their own learning process and construct their own knowledge. Moreover, Vygotsky (1978) believed an individual’s knowledge construction only occurs in a social and cultural environment, where one has to interact consistently with others to construct one’s knowledge.

Constructivist learning environments also demand a meaningful and authentic context for social and collaborative activities (Neo & Neo, 2009) and learners play a more active role as learners as they contribute to developing their own knowledge. Within such constructivist framework, learning takes place as learners progressively differentiate concepts into more complex understandings and also reconciles abstract understanding with concepts acquired from experience (Kwon & Cifuentes, 2009). Wen et al. (2004) found good online constructivist learning environments challenged learners’ existing concepts when they had to use inquiry learning and reflective thinking.

Therefore, the present study utilized social constructivist theory as a learning strategy to promote peer interactions for peer groups in an ML environment. Learners can pose questions and reflection related to traffic problems, collaborate with peers, learn new knowledge, and then improve their traffic violation reflection.
Reflection

Reflection is an intellectual and affective activity in which individuals are engaged in exploring their experiences to reach new understandings and appreciations (Boud et al., 1985). Additionally, Soloman (1987) viewed reflection as a harmonic process of an individual’s past experiences, actions and lessons learned. It can help with constructing one’s knowledge and meaning. In learning processes, reflection plays a major role and is very helpful for promoting learning performance (McNamara et al., 2006). Consequently, learners can review, test, and modify their own ideas while engaging in reflective practice.

Additionally, Yukawa (2006) pointed out the reflection process consists of three stages: (1) returning to experience, (2) attending to feelings, and (3) reevaluating experience. In these activities, collaborative learning environment plays an important role in providing the potential for applying new teaching and learning strategies. More specifically, a collaborative environment could stimulate group members to collaboratively reflect (i.e., Co-reflection) on the group performance and reach a shared conclusion about reflection. Co-reflection is defined as “a collaborative critical thinking process involving cognitive and affective interactions between two or more individuals who explore their experiences in order to reach new inter subjective understandings and appreciations” (Yukawa, 2006; p. 206).

According to this view, the present study thus emphasizes the effect of ML on learners’ reflection levels and further explores the educational implications of traffic violation reflection.

Gender and age

From the viewpoint of traffic violations, research on adults has proposed gender and age differences in compliance with traffic rules (Chang & Yeh, 2007). Besides, research indicated males expect less negative outcomes concerning traffic violations than female (Parker et al., 1992). Moreover, previous study has shown male pedestrians violate more rules than female pedestrians do (Moyano Diaz, 2002).

On the other hand, regarding the gender and age issues for computer-mediated education, it has captured the interest of computer educators for decades. Several studies have examined the role of gender and age in computer-related attitudes and the use of newer networked technologies (Shashaani, 1997; Uzunboylu et al., 2009). The results indicated learners’ characteristics have been regarded as important factors to predict their intention and performance. However, recent studies suggest the gender and age gap regarding technology use is disappearing (Sieverding & Koch, 2009; Uzunboylu et al., 2009). Additionally, to date, whether gender and age play a role in ML and traffic violations have not been investigated.

Based on the above, the present study utilized gender and age as the personal characteristics to understand its importance in influencing the attitudes toward mobile technology in an ML environment and the level of traffic violation reflection.

Research purpose and questions

This study was to integrate mobile communication technologies and a global positioning system (GPS) to construct an instant, convenient report of the mobile network service system named the Mobile Traffic Violation Reporting System (MTVRS), to improve learners’ traffic violation reflection level. Our evaluation mainly focused on answering the following four questions: (1) What are the effects of ML on traffic violations? (2) Do the learners’ gender and age affect the usefulness of ML systems for traffic violations? (3) Does the traffic violation reflection level on the post-test increase more significantly than the pre-test? (4) What are the effects of learners’ gender and age in group activities while taking different distributions for the reflection levels about traffic violations?
System implementation

Overview of MTVRS

MTVRS applies both mobile devices and a website to support learners in improving traffic violation reflection for ML. The proposed system designed two interaction interfaces, including a PC and a mobile device with GPS, through associated technologies, as shown in Figure 1. Learners were able to upload photos, process searches, and view records of traffic violations via mobile device or PC. Further, MTVRS collects data about which objects were shared and added by individuals, as well as the events of traffic issues and comments from the ML activities. All learning resources such as posted photographs, questions, discussions, comments and assessments were well liked by the system and support mobile devices and PCs synchronously for reflective learning.

As to supporting the services for mobile device, Churchill and Churchill (2008) stated mobile technology provides five affordances. The following five perceived affordances of mobile technology were explicated in this study:

- **Capture tool**: The learners can observe traffic violation situations and use a mobile device with cameras to photograph and transmit pictures through a wireless network, as shown in Figure 2(c).
- **Representational tool**: The learners can post and create relevant traffic violations issues to reflect and discuss with one another, as shown in Figure 2(d).
- **Multimedia-access tool**: The learners can view the resources of traffic violation events using the photograph mode or electronic maps, as shown in Figure 2(e, f).
- **Connectivity tool**: The learners can share information and their own opinion or thinking about the traffic violation events through email as well as on the website, as shown in Figure 2(g).
- **Analytical tool**: The learners can give feedback and comment on each traffic violation situation through peer assessment, as shown in Figure 2(h).

In short, MTVRS allows learners to login to the system for taking a traffic violation picture and uploading it, as shown in Figure 2(a, b). Additionally, MTVRS applies GPS technology to provide the location related information of every occurrence regarding reporting events for learning activities via a GPS receiver. The system then presents an electronic map to the learners, allowing them to browse and assess traffic violations in the map area. Accordingly, learners can use these learning resources to reflect their prior experiences and discuss with their peers. In contrast to supporting the services for a website, the main functions are similar to the services for mobile device. For example, MTVRS was also designed to provide learners with several online tools such as a discussion forum, posing/sharing reflection content, peer assessment, and search services (see Figure 3).
Figure 2. The MTVRS interfaces: (a) login menu, (b) main menu, (c) taking picture, (d) discussion, (e) browsing picture, (f) electronic map, (g) email, and (h) peer assessment

Figure 3. A webpage of traffic violations presentation in the discussion forum
Methods

Before the experiment, a mobile communication experience survey was used to assess learners’ ML experience. According to the results, about 80% of learners did not have relevant mobile device experience in using mobile communication. It implies most learners had insufficient knowledge with regard to the ability and experience operating mobile devices. Thus, before the experiment, learners had to study how to use mobile device to correctly operate each tool in the proposed ML environment.

Additionally, to know whether the participants have basic knowledge of traffic violations, a driving license automobile or motorcycle survey was used to assess learner’s traffic violation understanding. This result revealed all participants have at least one driving license and implied they have basic knowledge of traffic violations. In the following, the experiment was designed to answer the research questions.

Participants

This study recruited 32 volunteer participants, consisting of 17 males and 15 females, who were attending the Department of Information Management, Taiwan. Seventeen learners were graduate students and fifteen were college students. The mean ages of college and graduate students were 18 and 23, respectively. There were two learners’ characteristics to be considered in designing this study. One was gender, and the other was age. In addition, each characteristic had two groups divided according to the learners who were the same gender or similar age level. Therefore, there were four groups (conditions) in this study (see Table 1).

Before the experiment, all participants indicated they were highly interested in using mobile devices to enhance their learning experiences. Additionally, to ensure that all participants seriously take the learning activities, all participants were asked to read and understand the instructions and the procedure of the experiment as well as to post reflection content according to the design of the learning activity and share comments with their peers.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Age</th>
<th>N</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>High age</td>
<td>9</td>
<td>M-H Group</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>Low age</td>
<td>8</td>
<td>M-L Group</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>High age</td>
<td>8</td>
<td>F-H Group</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>Low age</td>
<td>7</td>
<td>F-L Group</td>
</tr>
</tbody>
</table>

Design of the learning activity

In this study, the learning activities were designed based on social constructivist theory and co-reflection learning strategy. These learning activities stress a need to place learners in a mobile reflective learning environment in which they can engage in effortful interactions with real-life context, the proposed system, and peers to improve the reflection. With learning devices, wireless connections, and the MTVRS, the reflective learning activities are executable without constraints of time or location. More detailed reflective learning activities are described as follows.

The learning topic of this study was “what situations may cause a traffic violation for a person or a car? And how can we avoid such violation?” In the beginning of the learning activity, with regard to the same characteristics of gender and similar age level, the participants were assigned to four different groups. After grouping, each learner then read through the pretest traffic violations article and co-reflection with peers in their group to respond to all questions based on which condition he or she was allocated. During the experimental period, each group was asked to contribute the learning resources and helpful reflection content. Thus, the learners used a Smart Phone to photograph local subjects, including traffic violation events and situations from their real-life surroundings. After finishing a subject creation, the learners had to reflect the event they had constructed. Additionally, learners were asked to share their comments or suggestions with their peers in their group. More specifically, during the learning process, each participant was guided by the following steps to complete the reflective learning activities with the proposed system: (1) to read the instructions, (2) to explore the real-life situations related to the learning topic, (3) to post reflection
content, (4) to observe peer’s comments and reflection contents, and (5) to revise original reflection contents. Here, a screen snapshot of the learning system is shown in Figure 4.

![Screen snapshot of the learning system](image)

**Figure 4.** A screen snapshot of the learning system

Regarding how to effectively facilitate the learning activity with the proposed system, all participants were actively engaged in reflecting the local traffic problems. Besides, they could utilize MTVRS’ four main different functions: accessing, discovering, discussing, and sharing to assist individuals to promote, facilitate, and enhance collaboration and interaction in relevant traffic violations issues through mobile communication technologies for synchronization. On the other hand, the learners could also review comments and photographs posted on a PC through the web page to share peer’s reflection contents synchronously.

In fact, these learning activities could be seen as social collaborative learning. Providing individual’s reflection contents and sharing their comments with peers could help learners improve their critical thinking abilities. Besides, the instructor also participated in these learning activities to enhance the learning process. According to the concept of cognitive conflict (Piaget, 1970), learners will try to revise their original reflection content to a higher level to avoid the unbalancing effect if they review others’ high-quality reflection content. Moreover, when a high-quality reflection content is posted by a learner, the instructor will also give positive reinforcement to the learner and guide other learners to look at this particular content.

**Instruments**

**Usefulness of mobile learning systems**

The usefulness of mobile learning systems (UMLS), a questionnaire adapted by Motiwalla (2007) and then Uzunboylu et al. (2009) was added to for use in measuring learners’ satisfaction with ML systems. The questionnaire mainly addresses three dimensions: asynchronous communication, synchronous communication, and mobile communication. The present study thus adapted this 23-item questionnaire (see Table 2), focusing on the usefulness of mobile devices for examining the effect of mobile learning systems (MLS) on traffic violations in an ML environment to collect data. For this measure, each item was evaluated on a five-point Likert scale (one is strongly disagree and five is strongly agree).

**Reflection levels**

Bain et al. (1999) classified reflection into five different levels: reporting, responding, relating, reasoning, and reconstructing, according to learners’ reflection content. However, in fact, the reflection content from learners was
too simple, incomplete, or they even misunderstood the questions. Therefore, the questionnaire could not satisfy the lowest level of reporting. To avoid this situation, this study adopted the modified evaluation criteria by Chen et al. (2009) to score the group’s reflection content of traffic violations article from each learning group (see Appendix 1).

**Procedures**

The designed learning activity was deployed for about four weeks including pretest stage, reflective learning activity and evaluation stage. Figure 5 shows the procedure of the experiment. In the first stage, the teacher introduced the learning activity and instructed the use of the proposed system which includes the idea of how to achieve reflective learning in the mobile setting. In this stage, a pre-test was administered to evaluate the initial reflection level of each group. In the second stage, according to the design of the learning activity mentioned above, the learners collaborated to complete the learning activity and contributed the reflection content and comments for the learning topic. After conducting the learning activities, the learners were asked to take a questionnaire regarding the usefulness of the proposed system and a post-test in the third stage. Further, all the learners in each group were interviewed to obtain their opinions about the present study.

![Figure 5. Experimental procedure](image)

**Results**

This section discusses results concerning: the UMLS questionnaire survey of ML activities based on different gender and age level, and the traffic violation reflection levels of the ML activities based on different groups.

**Effects of mobile learning on traffic violations**

To examine the effects of ML on traffic violations after using the proposed ML environment, SPSS statistic software was used to analyze the results of the UMLS questionnaire. After the learners completed the survey, the reliabilities
of the questionnaire (using Cronbach’s alpha) was 0.78. Analysis of the herein-considered sample showed a reasonable level of reliability (alpha=0.70). On average, learners reported highly positive attitudes towards MTVRS as a mobile reflective learning environment (M=4.01, SD=0.68).

Besides, Table 2 presents the descriptive statistics for each UMLS item. On the questionnaire survey, a mean of 4 or more was recorded for items 1, 3, 4, 6, 9, 12, 15, 17, 18, 21, 22 and 23. This result indicated most learners thought using the MTVRS could enhance their attitudes toward mobile technologies and be useful in better understanding traffic violations. For instance, item 1: “MLS is useful for creating traffic violation issues”, most learners agreed that MLS is useful for posing traffic violations issues (M=4.13, SD=0.49).

Following the use of mobile technologies with MTVRS, learners highly agreed with the benefits regarding their understanding of traffic violations problems from their real life surroundings. The highest score was recorded on the questionnaire for item 15: “MLS is an effective method for providing personalized information with regard to traffic violation issues” (M=4.38, SD=0.71). This result indicated mobile devices can provide scaffolding to the learners with effective approaches to investigation.

Further, in terms of item “Overall, satisfaction with MLS with regard to traffic violation issues is acceptable”, the result indicated learners generally accepted using MTVRS in an ML environment with regard to traffic violations was supported and satisfactory (M=4.16, SD=0.68). According to the result of item 23: “Using MLS gives users opportunity to join reflection about traffic violation issues without the limitation of time and space”, the questionnaire mean was 4.06 (SD=0.84). The result indicated learners had high agreement in using MTVRS to reflect traffic violation problems in the ML environment. In sum, the findings of the questionnaire were positive and satisfactory.

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MLS is useful for creating traffic violations issues.</td>
<td>4.13</td>
<td>0.49</td>
</tr>
<tr>
<td>2. MLS is a good reflection tool for creating traffic violations issues.</td>
<td>3.88</td>
<td>0.61</td>
</tr>
<tr>
<td>3. MLS is easy to use for creating traffic violations issues.</td>
<td>4.19</td>
<td>0.69</td>
</tr>
<tr>
<td>4. MLS made it easier to understand traffic violations issues.</td>
<td>4.09</td>
<td>0.64</td>
</tr>
<tr>
<td>5. MLS makes a good forum for interaction.</td>
<td>3.88</td>
<td>0.66</td>
</tr>
<tr>
<td>6. MLS makes it easier to reflect traffic violations problems with other learners.</td>
<td>4.13</td>
<td>0.61</td>
</tr>
<tr>
<td>7. MLS makes it easier to reflect traffic violations problems with the instructor.</td>
<td>3.97</td>
<td>0.78</td>
</tr>
<tr>
<td>8. MLS is a convenient platform to access reflection contents about traffic violations problems.</td>
<td>3.94</td>
<td>0.67</td>
</tr>
<tr>
<td>9. Overall, satisfaction with MLS with regard to traffic violations issues is acceptable.</td>
<td>4.16</td>
<td>0.68</td>
</tr>
<tr>
<td>10. MLS has the potential to become a good learning tool with regard to traffic violations issues.</td>
<td>3.78</td>
<td>0.71</td>
</tr>
<tr>
<td>11. MLS adds value to e-learning traffic violations issues.</td>
<td>3.84</td>
<td>0.68</td>
</tr>
<tr>
<td>12. MLS allows instant access to traffic violations issues regardless of your location.</td>
<td>4.00</td>
<td>0.62</td>
</tr>
<tr>
<td>13. MLS is useful as a supplement to traffic violations issues.</td>
<td>3.94</td>
<td>0.72</td>
</tr>
<tr>
<td>14. MLS is an effective learning aid or assistant for learners with regard to traffic violations issues.</td>
<td>3.88</td>
<td>0.66</td>
</tr>
<tr>
<td>15. MLS is an effective method for providing personalized information with regard to traffic violations issues.</td>
<td>4.38</td>
<td>0.71</td>
</tr>
<tr>
<td>16. MLS allows the conversion of idle time into productive time with regard to traffic violations issues.</td>
<td>3.81</td>
<td>0.59</td>
</tr>
<tr>
<td>17. MLS allows convenient access to reflection contents related to traffic violations problems– anywhere and anytime.</td>
<td>4.34</td>
<td>0.87</td>
</tr>
<tr>
<td>18. Information sent by MLS is more effective.</td>
<td>4.03</td>
<td>0.65</td>
</tr>
<tr>
<td>19. MLS allows access to information from the website about traffic violations issues.</td>
<td>3.94</td>
<td>0.50</td>
</tr>
<tr>
<td>20. MLS can be used as a supplemental tool for any existing course.</td>
<td>3.75</td>
<td>0.72</td>
</tr>
<tr>
<td>21. Information regarding traffic violations issues sent by MLS is more effective.</td>
<td>4.00</td>
<td>0.84</td>
</tr>
<tr>
<td>22. Information regarding traffic violations issues shared via discussing with MLS is more effective.</td>
<td>4.19</td>
<td>0.82</td>
</tr>
<tr>
<td>23. Using MLS gives users opportunity to join reflection about traffic violations problems without the limitation of time and space.</td>
<td>4.06</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Total feedback | 4.01 | 0.68 |
Comparisons of UMLS Questionnaire in terms of gender and age

To further examine the attitudes toward mobile technologies and traffic violations in ML for each gender and age level, the collected data were further divided into the same characteristic of learners by gender and age level. Table 3 and Table 4 showed a t-test of the independent sample, indicating a significant difference in UMLS scores between genders and ages by the end of the experiment. That is, the mean score of females on UMLS was 97.00 (SD=6.14) compared to 88.12 (SD=3.67) for males, a difference that was significant (t=-5.03, p=.000<.05). Besides, the mean score of college students (low age) on UMLS was 96.13 (SD=7.15) compared to 88.88 (SD=3.85) for graduate students (high age), a difference that was significant (t=-3.63, p=.001<.05).

In other words, learners used the MTVRS in the ML environment, the females’ mean score in terms of the UMLS questionnaire was significantly higher than the males’ mean score and the college students’ mean score in terms of the UMLS questionnaire was significantly higher than the graduate students’ mean score by the end of the test.

To further understand possible gender and age differences in the reflection level for traffic violations in the proposed ML environment, the reflection contents of the traffic violations among each group were examined as follows.

Examining the reflection levels of the traffic violations among groups

Descriptive statistics collected from the mean scores, standard error, and mean differences in each group on the traffic violation reflection levels between the pre-test and post-test are shown in Table 5. The pre-test mean scores of the reflection levels were greater than 4 in the groups with the M-H group (M=4.67, SD=0.50) and F-H group (M=4.78, SD=0.66). Clearly, their reflection levels were only higher than the reported level regarding the evaluation criteria for reflection levels (refer to Appendix 1). In other words, the learners’ reflection levels were generally lower than the reported level in the other groups. Based on this evaluation, the results of the reflection levels show learners had limited reflective performances when facing the present traffic violations questions at the beginning of the test.

Further, a one-way ANOVA found significant differences (F=13.79, p=.002<.05) in the mean scores on the group’s traffic violation reflection levels among the four groups during the pre-test stage and also found no significant differences (F=1.23, p=.360>.05) during the post-test stage (see Table 6). Scheffe’s post hoc test was used to further identify the source of significant F ratios in the pre-test (see Table 7). As can be seen, regardless of gender, the traffic violation reflection levels of the graduate student groups were significantly higher than those of the college student groups.
On the whole, the learners’ traffic violation reflection level improved by the end of the test. The mean differences between the pre-test and post-test positively increased among the four groups. Also, by the end of the experiment, analysis of the traffic violation reflection levels for all groups indicated the proposed ML environment significantly reduced the gap in reflective levels between the graduate student group and college student group. In particular, the female and college age groups (Group F-L) showed the greatest improvement, and the male and college age groups (Group M-L) were second. These findings indirectly explain why the analysis of the UMLS questionnaire score from females was significantly higher than males and the score from college students was significantly higher than graduates.

| Table 6. One-way ANOVA analyses of pre-test and post-test means on reflection level among groups |
| --- | --- | --- | --- | --- |
| | Sum of squares | df | Mean square | $F$ | Sig. |
| Pre-test | | | | | |
| Between groups | 5.38 | 3 | 1.79 | 13.79 | .002 * |
| Within groups | 1.04 | 8 | 0.13 | | |
| Total | 6.42 | 11 | | | |
| Post-test | | | | | |
| Between groups | 0.51 | 3 | 0.17 | 1.23 | .360 |
| Within groups | 1.11 | 8 | 0.13 | | |
| Total | 1.63 | 11 | | | |

$p < .05$

Therefore, based on the evidence, the findings indicated such an ML system could improve learners’ traffic violation reflection level. That is, the ML environment is extremely suitable for learners, regardless of gender or age. Further, this study interviewed all the participants to gain more detailed feedback. The feedback from all of the 32 learners showed their interest in the innovative approach. Specifically, the implications derived from the findings for the proposed study suggest learners’ understanding of traffic violations improved from using the MTVRS and the mobile devices.

**Discussion and conclusion**

This research has described an ML system providing learners with the methods to promote reflection levels for traffic violations. The MTVRS was designed and implemented to resolve problems possibly resulting from the use of an insufficiently didactic approach to teaching and learning about traffic education, as typically occurs in everyone.

Combining mobile technology with the concept of traffic violations, the MTVRS and its implementation model can help learners improve reflection levels of traffic violations through four functions: accessing, discovering, discussing, and sharing. With these applications of technological support, learners can frequently and effectively incorporate observable traffic violation situations into their feedbacks regarding contents, with provable benefits to reflection. Besides, according to the design of learning activity, learners learn to coordinate their efforts toward correcting and enhancing traffic violation knowledge and elaborate on relevant understanding in an ML setting, while observing learning subjects in the real world. These findings support the view of learning as a socio-cultural practice, as framed by Vygotsky (1978).

To examine the effects between traffic violations and ML as well as to further explore the learners’ traffic violation reflection level for the ML system in social constructivist theory, a questionnaire and evaluations of reflection levels were used to record the perceptions of learners in the study. The results revealed learners appreciated the use of ML for traffic violations and the use of the mobile technology and its many functions. Moreover, the results also revealed most learners’ perceptions of ML activities were very positive because the learners viewed their ML devices as an integral part of traffic violation learning. Additionally, most learners preferred ML activities to lectures on traffic violation issues because ML’s superior ability to promote communication, collaborative interaction, observation, and inquiry-based experiences.

Interestingly, regarding the roles of gender and age for the responses of the UMLS questionnaire, the results of this study indicated female learners achieved higher scores than male learners and college students achieved higher scores than graduate students. Perhaps females liked some of the specific applications concerning the traffic violation subject included in the present research more than males. The gender difference in the approach to technology
already described may play an important role here. Thus, the findings of the present study on learners’ characteristics in ML once again highlight the existence of gender differences in using technology.

In contrast to the effects of age differences in the experiment, graduate students’ feedback on the features of MTVRS in terms of perceived advantages and disadvantages were more than college students. Graduate students expressed viewpoints as to the needed improvements for our present study through the UMLS questionnaire responses, as a way of indirectly explaining why the analysis of their score was lower than college students.

Table 7. Post hoc multiple comparisons (Scheffe's method) for pre-test

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(I) Group</th>
<th>(J) Group</th>
<th>Mean difference (I - J)</th>
<th>SE</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
<th>Post Hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflection level</td>
<td>M-H</td>
<td>M-L</td>
<td>1.33*</td>
<td>0.29</td>
<td>0.013</td>
<td>0.30 - 2.36</td>
<td>M-H &gt; M-L</td>
</tr>
<tr>
<td></td>
<td>F-H</td>
<td>M-L</td>
<td>-0.11</td>
<td>0.29</td>
<td>0.984</td>
<td>-1.14 - 0.92</td>
<td>F-H &gt; M-L</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>M-L</td>
<td>1.22*</td>
<td>0.29</td>
<td>0.021</td>
<td>0.19 - 2.25</td>
<td>M-H &gt; F-L</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>M-H</td>
<td>-1.33*</td>
<td>0.29</td>
<td>0.013</td>
<td>-2.36 - 0.30</td>
<td>F-H &gt; M-L</td>
</tr>
<tr>
<td></td>
<td>F-H</td>
<td>F-L</td>
<td>-1.45*</td>
<td>0.29</td>
<td>0.008</td>
<td>-2.48 - 0.42</td>
<td>F-H &gt; F-L</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>F-H</td>
<td>-0.11</td>
<td>0.29</td>
<td>0.986</td>
<td>-1.14 - 0.92</td>
<td>F-H &gt; F-L</td>
</tr>
<tr>
<td></td>
<td>F-H</td>
<td>F-L</td>
<td>0.11</td>
<td>0.29</td>
<td>0.984</td>
<td>-0.92 - 1.14</td>
<td>F-H &gt; F-L</td>
</tr>
<tr>
<td></td>
<td>M-L</td>
<td>F-L</td>
<td>1.45*</td>
<td>0.29</td>
<td>0.008</td>
<td>0.42 - 2.48</td>
<td>F-H &gt; M-L</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>M-L</td>
<td>1.34*</td>
<td>0.29</td>
<td>0.013</td>
<td>0.31 - 2.37</td>
<td>F-H &gt; F-L</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>F-H</td>
<td>-1.22*</td>
<td>0.29</td>
<td>0.021</td>
<td>-2.25 - 0.19</td>
<td>M-L &gt; F-H</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>F-H</td>
<td>0.11</td>
<td>0.29</td>
<td>0.986</td>
<td>-0.92 - 1.14</td>
<td>M-L &gt; F-H</td>
</tr>
<tr>
<td></td>
<td>F-L</td>
<td>F-H</td>
<td>-1.34*</td>
<td>0.29</td>
<td>0.013</td>
<td>-2.37 - 0.31</td>
<td>M-L &gt; F-H</td>
</tr>
</tbody>
</table>

*p<.05

In fact, the use of ML and related technologies offers a means to address a broad range of topics and an avenue for learning. Previous research studies on how to enhance students’ mobile learning have largely focused on the role of the instructor as learning resource providers; however, there is relatively little research to investigate the effect of the role of the student as learning resource providers. From another learning perspective, educational resources could be provided by the learners within well-considered learning solutions for learning (using guided learning strategies and technology-enhanced learning approaches). Accordingly, learners could actively discover their learning contents for themselves rather than by passive guidance to achieve the purpose of knowledge acquisition and sharing.

Although several studies indicate significant support for ML, there seems to be no research examining the effect between ML and the traffic violation reflection level through integrating mobile devices with the web-based service system. The findings of this study give positive preliminary evidence that mobile technologies, the guided reflective learning activity and an outdoor inquiry-based strategy are useful tools in teaching learners about traffic violations.

Besides, mobile devices provided a means to engage males and females to develop positive attitudes toward traffic violations reflection, a finding supported by prior research (Uzunboyu et al., 2009), and the use of mobile technology and ML benefited both genders. Using mobile device outside of the classroom has excited and interested many participants in this study. They seemed comfortable with the technologies used and with the educational use for mobile devices. Long-term, ML will likely provide a suitable platform for use in a wide range of educational settings.

In the future, MTVRS may be able to generate and facilitate lifelong learning and communities of practice, by allowing users to share resources and experiences about traffic violations. The authors consider this kind of instruction can lead to good learning experience, and learners will feel the ML process is enjoyable and useful. Our next step will be to integrate different subjects and disciplines to apply these new tools in productive ways for learning.
Acknowledgements

This work was supported in part by the National Science Council (NSC), Taiwan, ROC, under Grant NSC 97-2815-C-150-031-E.

References


### Appendix 1: Evaluation criteria for reflection levels

<table>
<thead>
<tr>
<th>Reflection levels</th>
<th>Evaluation criteria</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsense</td>
<td>Reflection content made no sense or was missing</td>
<td>1</td>
</tr>
<tr>
<td>Simple</td>
<td>Reflection content limited to yes or no</td>
<td>2</td>
</tr>
<tr>
<td>Incomplete</td>
<td>Reflection content incomplete</td>
<td>3</td>
</tr>
<tr>
<td>Reporting</td>
<td>Only repeated content already contained in the article</td>
<td>4</td>
</tr>
<tr>
<td>Responding</td>
<td>(a) Only used a few concepts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) Stated observed phenomena, but with no reasoning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) Stated personal affection only</td>
<td></td>
</tr>
<tr>
<td>Relating</td>
<td>(a) Stated relationships within context</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>(b) Explained the cause</td>
<td></td>
</tr>
<tr>
<td>Reasoning</td>
<td>(a) Detailed explanation of the rationale</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>(b) Combined theory and practice</td>
<td></td>
</tr>
<tr>
<td>Reconstructing</td>
<td>(a) Expressed higher order of reconstruction</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>(b) Organized theories, rules and experiences using a systematic approach</td>
<td></td>
</tr>
</tbody>
</table>

(Adapted from Chen et al., 2009)
The Comparison of the Opinions of the University Students on the Usage of Blog and Wiki for Their Courses

Ummuhan Avci and Petek Askar
Faculty of Education Department of Computer Education and Instructional Technology, Baskent University, Ankara, Turkey / 1Faculty of Arts and Sciences Department of Sociology, Izmir University of Economics, Izmir, Turkey // uavci@baskent.edu.tr // petek.askar@ieu.edu.tr

(Submitted August 26, 2010; Revised March 21, 2011; Accepted May 10, 2011)

ABSTRACT
The purpose of this study was to investigate the use of blogs and wikis as constructive tools in the computer courses of prospective teachers and compare them with respect to perceived usefulness, perceived ease of use, intention, self efficacy, and anxiety. Ninety-two students who were enrolled in various teacher education programs used blog and wiki for their courses. After their experiences with blog and wiki, the data were collected by administrating the instrument developed by the researchers. The results showed that students were positive to blog and wiki usage in the teaching-learning process. However they found wiki more useful. Both perceived usefulness and self efficacy variables explain 71% of blog and wiki usage. This value points to a very high and significant correlation. Perceived usefulness has been identified as the variable that could explain intention by itself at the most.

Keywords
Unified theory of acceptance and use of technology (UTAUT), Blog, Wiki

Introduction
In the recent years, the rapid changes of information and communication technologies have brought the innovations in communication and information sharing. These innovations have contributed to further enhance the teaching and learning process in higher education. Time and space problems have eliminated through these changes and communication, interaction and information sharing between individuals have also become easier.

Along with the emergence of the virtual classrooms concept, the prevalence of distance education and increasing use of electronic elements in the traditional classroom, educators are realizing that the interaction have gained importance. Therefore, the new studies that are commonly called social software are needed to improve the student-teacher and student-student interaction. Social software supports the desire of individuals to be pulled into groups to achieve goals (Connell, 2006). Social software refers to applications such as wikis, blogs and social networks, used for sharing multimedia, audio or visual content, as well as text. Social software tools support learning in a variety of ways: sharing of resources, collaborative learning and peer-to-peer learning, also facilitate communication and collaboration between participants by providing awareness. Also it is possible to share the information resources of other users and get feedback. In addition each member’s contribution to the work within the group could be shared (Anderson, 2005; Minocha, 2008; Conole & Culver, 2010).

Web 2.0 technologies, such as blog, wiki, podcast, facebook, flickr, RSS are called social software tools in respect to their features. Web 2.0 is a set of internet services which encourage internet users to participate in various communities of knowledge building and knowledge sharing. Web 2.0 also better supports group interaction and fosters a greater sense of community. Web 2.0 encourages more active learning and enables feedback from tutors to learners, this tutor-student interaction leads to further increasing of student motivation (Crook, 2008). Minocha (2008) stated that Web 2.0 seems to match well with modern thinking on educational practice. In particular, it promises learners of new opportunities to be independent in their study and research. Web 2.0 technologies encourage a wider range of expressive capability. They facilitate more collaborative ways of working.

Web 2.0 technologies are part of the changes in information and communication technology. Web literacy improves the base of individuals’ skills and it can reach a wider range of users with these applications. Blog and wiki applications improve the teacher and learner’s social interaction, their creativity, their ability to express themselves and their high-level thinking skills. They also enable them to configure the information. At the same time weblog and wiki applications allow users to communicate synchronously and work collaboratively. Students can participate
actively in the learning process. Because of these student-centered new technologies, students can work at their own pace and at any time. These applications provide the opportunity to re-use always and everywhere. They also allow peer evaluations (Godwin-Jones, 2003; Huffaker, 2005; Aronsson, 2002).

Many instructors and learners start using Web 2.0 technologies for effective teaching and learning process. The social aspects of Web 2.0 technologies provide a framework for collaborative group learning. The possibilities of Web 2.0 environments, which enable learners to work with evolving content online through reading, writing, editing and communicating require teachers to reshape their approaches. However, according to Dron (2007) there are many ways that social software can fail to address the needs of learners. Therefore he assessed the potential of social software (blogs, wikis, social networking) in educational contexts and focused on the use of these tools for learners. He identifies a set of highly interconnected principles for educational social software that meets the needs of learners. These are: the principle of adaptability (compatibility); the principle of evolvability (unfixed systems); the principle of parcellation (connections between systems should emerge and not be prescribed); the principle of trust (goodwill); the principle of stigmergy (using signs to guide, not constrain); the principle of context (awareness of virtual ecosystems); the principle of constraint (awareness of what is excluded); the principle of sociability; the principle of connectivity (interconnectedness); the principle of scale (where small iterations underpin larger ones). According to the writer, it is likely that social software will be successful in self-organizing for the benefit of learners with adherence to each of these principles.

**Blogs and wikis**

Blogs (abbreviated from weblogs) are personal web pages which are easy to use. Blogs give opportunity to people to present information on various topics and interact with others without requiring the design information (Sim & Hew, 2010). In the blog platforms, students can follow their own learning process and can log about their reading, learning and working process in their lessons. Students can use blogs to get information from each other or to learn topics which they do not understand. They can also debate on any topics (Namwar & Rastgoo, 2008).

Wiki is a platform that individual groups can work together synchronously on an idea. Wikis refer to collaborative websites that allow users to interact by adding, removing, or editing site content (Mindel & Verma, 2006). Wiki shows structural similarities with the blog page. Even though the blog has a single author, everyone is a writer on wiki pages. On the blog, visitors can send their comments to the author’s message, but they don’t have the right to change the existing content. However, on the wiki, they can change all the text by page editing options (Altun, 2005). Wikis are useful in educational settings in which they support individual learning, more socially defined search structures and promotion of collaboration through group editing and peer review is supported. They also provide an attainable store for resources (Snelling & Karanicolas, 2008).

**UTAUT: Unified theory of acceptance and use of technology**

Unified Theory of Acceptance and Use of Technology (UTAUT) was created by Venkatesh, Morris, Davis and Davis (2003). They reviewed large number of theories along with models and selected the appropriate structures in these theories and models. In their study the researchers reviewed user acceptance literature and discussed eight prominent models, compared the eight models and their extensions, formulated a unified model that integrates elements across the eight models, and empirically validated the unified model. The purpose of this unified model is to identify the behavioral intention putting the factors such as the gender, age, experience, and voluntariness of use among the variables such as performance expectancy, effort expectancy, social influence, and facilitating conditions which affect user acceptance and usage (Venkatesh et al., 2003).

The researches are based on this model observed that the previous models explain 40% of users’ acceptance of the technology, UTAUT explains approximately 70% of users’ acceptance of the technology. Venkatesh et al. (2003) stated that UTAUT explains as much as 70% of the variance of usage intention with the long-term studies. Researchers find three variables which are performance expectancy, effort expectancy, and social influence have an effect on technology usage intention with this model.
Literature review

The related variables of this research (UTAUT, blog, wiki) and the studies which examined the relationship between these variables are included in this section of the research.

Web 2.0 social software tools, blog and wiki

Minocha (2008) examined the use of social software in enhancing student learning and engagement. She found that social software tools support a variety of ways of learning: sharing of resources, collaborative learning, problem-based and inquiry-based learning, reflective learning and peer-to-peer learning. Students gain transferable skills of team working, negotiation and communication, individual and group reflection, and managing digital identities. Ajjan and Hartshorne (2008) researched the university students’ adaptation of Web 2.0 tools. In their study, findings indicated that some faculty members feel that some Web 2.0 technologies could improve students' learning, their interaction with faculty and with other peers, their writing abilities, and their satisfaction with the course.

The review of the Web 2.0 researches, showed that Web 2.0 tools provide an active participation to the process of learning and can be used effectively in the individual and collaborative learning environments. Therefore there is a need to investigate the opinions about the usage of Web 2.0 tools. In addition, blog and wiki has come more to the forefront in the researches about Web 2.0 tools.

Masek and Hingston (2007) examined the use of wikis and blogs in aiding group work for a large and diverse student-based team. In this study, students using these tools for the completion of assignments gave a high rate for the flexibility to work anywhere with an internet connection. Data was obtained by keeping log statistics and questionnaires which were coordinated for receiving feedback about the blog and wiki tools. The results suggested that the flexibility of these tools provides more motivation for students. While blogs are time based, with separate chronologically ordered entries not designed for future editing, wikis are completely free in their formats.

Williams and Jacobs (2004) indicated that blogs are personal online diaries. Their research explored the potential of blogs as learning spaces for students in the higher education sector. Their research showed that blogs have the potential to be a transformational technology for teaching and learning, also provide more communication and interaction between students. According to Yang (2009) blog is a platform that provides students’ critically reflect on their learning processes and help their professional growth. Blog was used by the participants as a discussion space in this study. Participants can get more opportunities to make comments, challenge each other’s point of view and learn from each other by using blog platform.

According to Su and Beaumont (2008) wiki technology has enabled a new form of online communication. Moreover wikis can help provide an efficient, flexible, user friendly and costeffective interface for collaboration, knowledge creation and archiving, and student interaction. Wikis provide freedom of authoring and editing for any user. In their research, students constructed literature reviews in the wiki, with peer and tutor review of contributions. The analysis of their study showed that the wiki learning environment has enhanced students' learning, provide quick feedback from tutors, other users and active peer interaction. Majchrzak, Wagner and Yates (2006) stated that users attained three main types of benefits from wikis: enhanced reputation, work made easier, and helping the organization to improve its processes. Users could be categorized as “synthesizers” and “adders” in accordance with their variety of contributions. Synthesizers’ frequency of contribution was affected more by their impact on other wiki users, while adders’ contribution frequency was affected more by being able to accomplish their immediate work. Also they indicated that synthesis of the knowledge is as important as adding the knowledge.

According to the studies carried out in this field, it can be said that flexibility of using wiki and blog increases the users’ interest and motivation. Wikis have an active role in the configuration of information process, they support the collaborative learning environment and they are mostly preferred for cooperative works. Blogs are the critical reflection platforms where the students can often feel themselves relaxed and also make contact with each other.
Blog, wiki and unified theory of acceptance and use of technology

Koca and Usluel (2007) examined teachers’ usage of information and communication technology as reflected in Venkatesh et al. (2003) variables of UTAUT. They also investigated teachers’ perceptions of information and communication technology (ICT) in relation to their intention to use computer. The purpose of their study was to focus on more effective ways to use ICT in schools. "Information and Communication Technology Acceptance and Usage Scale" was developed for data collection. Principal component analysis was used for construct validity of the scale. According to the principal component analysis results, 34 items were gathered under 8 factors. These were 1) intention 2) perceived usefulness 3) perceived ease of use 4) anxiety 5) social impact 6) self efficacy 7) facilitating conditions 8) voluntariness. These eight factors explained 73.73% of test scores. The score of Cronbach’s $\alpha$ coefficient was calculated as 0.91. Teachers’ perceived usefulness, perceived ease of use, social factors, and self efficiency have significant effects on intention to use ICT.

Huang, Yoo and Choi (2008) examined the relationship between the level of using Web 2.0 tools (blog, wiki, facebook, youtube, skype, second life) and learning styles of university students. Researchers developed a measurement tool according to the UTAUT in order to determine the level of using Web 2.0 tools. At the end of the research many students found Web 2.0 tools easy to use. Liu (2006) investigated the wiki’s future to understand its development. The research was based on investigation of behavior of wiki users and provided useful suggestion for future research and practical usage. The research was based on UTAUT. The researcher investigated wiki users’ acceptance behavior, through valid questionnaires. The results showed that performance expectancy, effort expectancy, facilitating conditions, and user involvement have positive effects on wiki users’ intention of use. Social influences have no significant effect on wiki users. User involvement has the most significant effect on wiki users’ intention of use. The researcher explained that the results might be utilized for future wiki system development suggestions to improve on its usability and adaptability.

The results of the researches showed that unified model explains the usage of technology intention substantially. It is observed that perceived usefulness and perceived ease of use come to the fore in related studies with UTAUT.

Research questions

Nowadays, there is an increasing interest in using blog and wiki, in teaching-learning process. Despite having great benefits in many ways, the differences between blog and wiki need to be investigated. Researches have shown that while the blog has been used for more personal purposes, the wiki has been used more for group purposes (Gunawardena et al., 2009; Parker & Chao, 2007; Masek & Hingston, 2007; Godwin-Jones, 2003). Differences between blog and wiki also reveal of their usage. Therefore, a comparative review of these two new social software tools is important with respect to perceived usefulness, perceived ease of use, intention, self-efficacy, and anxiety. Moreover, this research includes the relationship between blog and wiki usage intention.

This is one of the few researches which analyses together these two important and new social software tools: blog and wiki. The research is also very important for providing a comparative perspective about these tools. In this study the research questions were:

i. Is there a difference between blog and wiki in terms of perceived usefulness, perceived ease of use, intention, anxiety and self efficacy?

ii. Do perceived usefulness, perceived ease of use, anxiety, and self-efficacy variables predict the blog’s usage intention in a significant way?

iii. Do perceived usefulness, perceived ease of use, anxiety, and self-efficacy variables predict the wiki’s usage intention in a significant way?

Methodology

Research design

This study compared blog and wiki in terms of perceived usefulness, perceived ease of use, intention, self efficacy, anxiety and examined the intentions of usage. Casual-comparative research model was used in this study. Data were
collected through Personal Information Survey, Blog Acceptance and Usage Scale and Wiki Acceptance and Usage Scale.

Participants

Data collection process of the study was carried out in two stages. The first stage is the validity and reliability studies of the data collection tools, namely, "Wiki Acceptance and Usage Scale" and "Blog Acceptance and Usage Scale". The validity and reliability analysis of the scales were performed on a sample of 77 blog users and 51 wiki users who had used these tools in their lessons before.

After the validity and reliability studies of measuring tools, the research group was determined of 92 students from departments of Turkish Language Education, Elementary Mathematics Education, Early Childhood Education and Mixed Group students of Foundation University. Thirty students from departments of Turkish Language Education, 26 students from departments of Elementary Mathematics Education, 25 students from departments of Early Childhood Education and 11 students from departments of Mixed Group participated to the research.

Thirteen percent of the participants men, 87 percent women in a total of 92 students participated to the research. Demographic characteristics of the research group were examined by looking at their computer usage experience, duration of computer usage and blog-wiki usage status. Nine percent of the students stated that they were beginners, 87% said that they were at the intermediate level, and 4% said that they were advanced users. In response to a question as to how long they had been using computer, 3% said less than 1 year, 14% reported 1–3 years, 24% said 4–6 years, and 59% said more than 6 years. Seventy five students expressed that they were not able to use blog and 80 students expressed that they were not able to use wiki before.

Study procedure

First of all, blog and wiki platforms have been installed on the server of the university under the wordpress application.

In Computer II course blog applications were completed in two weeks and wiki applications were completed in three weeks. Between blog and wiki applications, two weeks break was given. Implementation process was carried out in parallel at all departments.

First, the trial platforms were created for the students to provide equal knowledge level and create basic knowledge about blog and wiki. A few implementations were realized on these trial platforms. Then in the blog platform, the students discussed on the subjects of "Positive and negative effects of computers and internet on children and young people" and "Advantages and disadvantages of computer based education". All classes discussed amongst themselves. However, all discussions were followed by individual students as well. Thus, a small discussion platform was created for the Faculty of Education. After the blog application, all classes selected a unit title within their own courses and created content on wiki platform by working in groups of 3–4 persons. However, all content creation works were followed by students as a group. Thus, a small encyclopedia was created for the Faculty of Education.

The process of discussion on blog platform lasted two weeks and the process of content creation on wiki platform lasted three weeks. The scales administered after the treatment.

Instruments

Personal Information Survey

Scale includes a total of nine questions about the gender of participants, their departments, their classes, their computer using time, and their computer using experience. Personal information form was applied to the students before the other scales.
7-point Likert-type scales (Blog Acceptance and Usage Scale, Wiki Acceptance and Usage Scale) ranging from 1 (strongly disagree) to 7 (strongly agree) were used in this study. These scales were adapted from the “Internet and Communication Technologies Acceptance and Usage Scale” which was developed by Koca and Usluel (2007), based on the Technology Acceptance and Usage Model which was developed by Venkatesh et al. (2003).

For the reliability evidences of the scales and their dimensions, Cronbach’s α coefficients were calculated which can be seen in Table 1 and Table 2. The coefficients for Blog Acceptance and Usage Scale’s each dimensions were 0.82 for perceived usefulness, 0.81 for perceived ease of use, 0.86 for intention, 0.79 for anxiety, 0.50 for social effect, 0.58 for self-efficacy, 0.45 for facilitating conditions (Table 1) and for Wiki Acceptance and Usage Scale’s each dimensions was 0.80 for perceived usefulness, 0.85 for perceived ease of use, 0.80 for intention, 0.70 for anxiety, 0.50 for social effect, 0.71 for self-efficacy, 0.49 for facilitating conditions (Table 2).

As shown on the tables, “Social Effect” and “Facilitating Conditions” dimensions were kept outside of data analysis, because their α values are smaller than 0.60 in both scales.

### Data analysis

In data analysis section descriptive statistics, dependent samples t-test and stepwise multiple regression analysis were used. In this study, the level of significance was taken as 0.01 because of multiple t-tests.

### Findings

The statistics of scores which were calculated for each variable of “Blog Acceptance and Usage Scale” and “Wiki Acceptance and Usage Scale” are shown at Table 3.
Comparison of blog and wiki

"Is there a difference between blog and wiki in terms of perceived usefulness, perceived ease of use, intention, anxiety and self-efficacy?"

Dependent samples t-test was used to determine whether there is a difference between blog and wiki in terms of perceived usefulness, perceived ease of use, intention, anxiety and self-efficacy (Table 4).

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>( \bar{x} )</th>
<th>SS</th>
<th>t-test</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness</td>
<td>92</td>
<td>24.09</td>
<td>5.753</td>
<td>91 5.411</td>
<td>.000 .563</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>92</td>
<td>27.23</td>
<td>5.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>92</td>
<td>21.09</td>
<td>4.729</td>
<td>91 2.121</td>
<td>.037 .220</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>92</td>
<td>22.14</td>
<td>4.349</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention</td>
<td>92</td>
<td>24.09</td>
<td>6.187</td>
<td>91 2.355</td>
<td>.021 .244</td>
</tr>
<tr>
<td>Intention</td>
<td>92</td>
<td>25.40</td>
<td>5.239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>92</td>
<td>8.28</td>
<td>4.512</td>
<td>91 3.233</td>
<td>.002 .337</td>
</tr>
<tr>
<td>Anxiety</td>
<td>92</td>
<td>9.75</td>
<td>4.359</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>92</td>
<td>13.37</td>
<td>3.726</td>
<td>91 2.122</td>
<td>.037 .220</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>92</td>
<td>14.26</td>
<td>3.905</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The difference between the scores of blog and wiki’s perceived usefulness variable was found significant (\( p < .001 \)). The average of wiki’s perceived usefulness scores was calculated as 27.23, while the average of blog’s perceived usefulness scores was 24.09. Effect size of the study (Cohen's \( d \)) was .563. This value indicated a medium level of effect size according to the Cohen's Standard. It can be confirmed that the students found wiki more useful than blog.

The difference between the scores of blog and wiki’s perceived ease of use variable was not found significant (\( p = .037 \)). The average of wiki’s perceived ease of use scores was calculated as 22.14, while the average of blog’s perceived ease of use scores was 21.09. Effect size of the study (Cohen's \( d \)) was found .220.

The difference between the scores of blog and wiki’s intention variable was not found significant (\( p = .021 \)). The average of wiki’s intention scores was 25.40, while the average of blog’s intention scores was 24.09. Effect size of the study (Cohen's \( d \)) was .244.

The difference between the scores of blog and wiki’s anxiety variable was significant (\( p = .002 \)). The average of wiki’s anxiety scores was 8.28, while the average of blog’s anxiety scores was 9.75. Effect size of the study (Cohen's \( d \)) was .337. This value indicated a small level of effect size according to the Cohen's Standard. According to the obtained findings, it can be stated that although the difference between the scores of blog and wiki’s anxiety variable was found significant, the anxiety level of both implementation are low.

The difference between the scores of blog and wiki’s self-efficacy variable was not found significant (\( p = .037 \)). The average of wiki’s self-efficacy scores was 14.26, while the average of blog’s self-efficacy scores was 13.37.
Prediction of blog’s usage intention

“Are perceived usefulness, perceived ease of use, anxiety and self-efficacy variables predicting the blog’s usage intention in a significant way?”

In this problem, intention was taken as the dependent variable, while perceived usefulness, perceived ease of use, anxiety and self-efficacy were taken as independent variables of Blog Acceptance and Usage Scale. Then multiple regression analysis was utilized. Correlation coefficients between variables are also given at Table 5.

Table 5. Correlation coefficients between variables

<table>
<thead>
<tr>
<th>Blog</th>
<th>Perceived usefulness</th>
<th>Perceived ease of use</th>
<th>Anxiety</th>
<th>Self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention</td>
<td>.789</td>
<td>.176</td>
<td>.115</td>
<td>.648</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>.127</td>
<td>.110</td>
<td>-.402</td>
<td>.505</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td></td>
<td></td>
<td>.566</td>
<td>.109</td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td>.566</td>
<td>.109</td>
</tr>
</tbody>
</table>

When examining the correlation coefficients between variables, the highest relationship found was between intention and perceived usefulness.

Stepwise model was used during the regression analysis. The results of regression analysis about the prediction of the intention of Blog Acceptance and Usage Scale variables are given in Table 6 and Table 7.

Table 6. The results of regression analysis for the prediction of the intention

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std. Error</th>
<th>Beta (β)</th>
<th>t</th>
<th>p</th>
<th>Zero order</th>
<th>Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Perceived usefulness</td>
<td>.849</td>
<td>.070</td>
<td>.789</td>
<td>12.182</td>
<td>.000</td>
<td>.789</td>
</tr>
<tr>
<td>2</td>
<td>Constant</td>
<td>.449</td>
<td>1.639</td>
<td>.274</td>
<td>.785</td>
<td>.000</td>
<td>.708</td>
</tr>
<tr>
<td>Intention</td>
<td>Perceived usefulness</td>
<td>.667</td>
<td>.071</td>
<td>.620</td>
<td>9.448</td>
<td>.000</td>
<td>.789</td>
</tr>
<tr>
<td></td>
<td>Self-efficacy</td>
<td>.566</td>
<td>.109</td>
<td>.341</td>
<td>5.193</td>
<td>.000</td>
<td>.648</td>
</tr>
</tbody>
</table>

Note. $R^2 = .710; F(2, 89) = 109.096; p<.001$

Table 7. The results of regression analysis for the prediction of the intention

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.622</td>
<td>.618</td>
<td>148.410</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>.710</td>
<td>.704</td>
<td>109.096</td>
<td>.000</td>
</tr>
</tbody>
</table>

Note. 1: Perceived usefulness; 2: Perceived usefulness + Self-efficacy

When the t-test results are examined in accordance to significance of regression coefficients, it is seen that perceived usefulness and self-efficacy are explanatory predictors of blog usage intention. According to the standardized regression coefficient $\beta$, the relative order of importance of this two predictor variables on the dependent variable: perceived usefulness and self-efficacy.

Perceived usefulness variable first entered into the model and gave significant results ($p<.001$). Perceived usefulness explains approximately 62% of intention by oneself. Both perceived usefulness and self efficacy variables give very high and significant correlation ($R^2 = .71, p < .001$). These two variables explain approximately 71% of blog usage together. Perceived usefulness has been determined as the highest variable that could explain intention by itself between Blog Acceptance and Usage Scale variables.

Prediction of wiki’s usage intention

“Are perceived usefulness, perceived ease of use, anxiety and self-efficacy variables predicting the wiki’s usage intention in a significant way?”
In this problem, intention was taken as the dependent variable, while perceived usefulness, perceived ease of use, anxiety and self-efficacy were taken as independent variables of Wiki Acceptance and Usage Scale. Then multiple regression analysis was utilized. Correlation coefficients between variables are also given at Table 8.

Table 8. Correlation coefficients between variables

<table>
<thead>
<tr>
<th>Wiki</th>
<th>Perceived usefulness</th>
<th>Perceived ease of use</th>
<th>Anxiety</th>
<th>Self-efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention</td>
<td>.817</td>
<td></td>
<td>-.097</td>
<td>.615</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>.374</td>
<td></td>
<td>-.060</td>
<td>.536</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td>-.373</td>
<td></td>
<td>.574</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td>-.234</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When examining the correlation coefficients between variables, the highest relationship found was between intention and perceived usefulness.

Stepwise model was used during the regression analysis. The results of regression analysis about the prediction of the intention of Wiki Acceptance and Usage Scale variables are given in Table 9 and Table 10.

Table 9. The results of regression analysis for the prediction of the intention

<table>
<thead>
<tr>
<th>Model</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.667</td>
<td>.664</td>
<td>180.504</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>.712</td>
<td>.705</td>
<td>109.769</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 10. The results of regression analysis for the prediction of the intention

The t-test results on the significance of regression coefficients showed that that perceived usefulness and self-efficacy are explanatory predictors of wiki usage intention. According to the standardized regression coefficient β, the relative order of importance of these two predictor variables on the dependent variable: perceived usefulness and self-efficacy.

Perceived usefulness variable first entered into the model and gave significant results (p<.001). Perceived usefulness explains approximately 67% of intention by itself. Both perceived usefulness and self-efficacy variables give very high and significant correlation (R² = .71, p < .001). These two variables explain approximately 71% of wiki usage together.

Results and discussion

The results of the study revealed that when the blog and wiki were compared in terms of perceived usefulness and anxiety, significant differences were found in favor of wiki. However, significant differences were not found when the blog and wiki were compared in terms of perceived ease of use, intention, and self-efficacy. The effect size of the perceived usefulness variable is higher than other variables.

The results showed that students were positive to blog and wiki and they found wiki more useful. The results about wiki are more positive probably because wiki is a flexible collaborative platform that can be used continuously. Students can change the content, correct their errors, create their own resources, and produce the material together. Also Mindel and Verma (2006) discovered that use of wikis in an educational setting follows belief in a social (learner-based) construction of knowledge, where learners collaborate and share in the learning process.
When the implementation process is examined, we can see that group work has been done on the wiki platform whereas individual study has been carried on the blog platform. Therefore group work has caused some differences between wiki and blog.

Wikis and blogs contribute to the students’ work positively in learning environments. In this study, while blog is used as a content discussion platform, wiki is used as a content development platform within the scope of computer course. Blogs and wikis are two important course content discussion, development and management technologies that enable an interactive engagement amongst students and between students and teachers. Therefore educational possibilities of blogs and wikis should be considered.

Usage of blogs can address some of the theoretical underpinnings that are summarized below (Glogoff, 2005):

- In instructional blogging, as a knowledge centered instructional tool, the instructor designs research activities that engage students in discussions with practitioners.
- In learner-centered blogging, the instructor gives positive feedback to students regarding their comments and by posting comments for discussion.
- For providing community-centered instruction, blogging supports the importance of social and peer interaction. As a receptive learning tool, blogging can encourage students to acquire information from resources and reflect on what they have gathered.
- In a directive learning environment, blogs provide students with equal access to information, to expand students’ understanding of specific issues, and to direct students to explore additional material.

Wikis are also very useful tools for educational purposes, since they encourage student participation and also a sense of group community. The educational benefits of wikis can be summarized as follows by Duffy and Bruns (2006):

- Wikis offer an online space for collaborative authorship and writing.
- Wikis are available online for all web users or for members of specific communities, and include version control tools that allows authors to track the history of specific pages, and the history of their personal contributions.
- Using wikis, students can easily create simple websites without prior knowledge or any skill in programming, thus eliminating the time overhead necessary to develop these skills.
- A wiki allows teachers and learners to see the evolution of a written task, and to continually comment on it, rather than offering comments only on the final draft.
- A wiki can also be very useful for tracking and streamlining group projects.

Venkatesh et al. (2003) have expressed in their study that perceived ease of use, perceived usefulness, social effect, and self efficacy have strong effects on intention. In this study, both perceived usefulness and self efficacy variables explain 71% of blog and wiki usage. This value points to a very high and significant correlation. Perceived usefulness has been determined as the highest variable that could explain intention by itself.

According to the students, wiki and blog make their tasks faster, make the process easier, increase their performance and productivity, so that they have more positive views and intentions to use blog and wiki in the future. Therefore perceived usefulness is thought to be the most determining factor. However, wiki’s perceived usefulness factor can explain intention by oneself more than blog’s perceived usefulness factor. In the implementation process, it can be said that creating content as a group of students have an impact on perceived usefulness.

Some studies has also supported this view. Davis, Bagozzi and Warshow (1989) explained in their studies that perceived usefulness is the most important determinant which influenced peoples’ intentions and perceived ease of use had a small but significant effect on intentions as well. Moreover Ma, Anderson and Streith (2005) investigated student teachers' perceptions of computer technology in relation to their intention to use computers. Their study results indicated that student teachers' perceived usefulness of computer technology had a direct significant effect on their intention to use it and student teachers' perceived ease of use had only an indirect significant effect on intention to use computer technology.

Even though the research group's technology experience was limited, most of the results were significant and positive. It can be said that these results are thought to be important for the future of this social software tools such as blog and wiki. Moreover, the usage of blog and wiki was found useful and easy by the students, also their anxiety about using them was very little, so it can be said that this social software tools will become widespread any longer.
References


Educational Affordances of a Ubiquitous Learning Environment in a Natural Science Course

Tan-Hsu Tan, Min-Sheng Lin, Yu-Ling Chu and Tsung-Yu Liu1*

Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan // 1Department of Multimedia and Game Science, Lunghwa University of Science and Technology, Taoyuan County, Taiwan // thtan@ntut.edu.tw // mslin@ee.ntut.edu.tw // chu_yuling@tp.edu.tw // joye.liu@msa.hinet.net

*Corresponding author

Submitted August 26, 2010; Revised December 11, 2010; Accepted April 15, 2011

ABSTRACT

Educational affordances are worthy of being explored because the affordances of Information and Communication Technologies (ICT) provide the pedagogical effects for promoting cognitive acceleration, increasing the self-management of students, facilitating data collection, and presentation in science learning. This study aims to explore educational affordances on a course with a learning environment. To address this research goal, this study employed a ubiquitous learning environment termed the Environment of Ubiquitous Learning with Educational Resources (EULER) and conducted a natural science course for eight weeks. The participants included elementary school teachers and students. Detailed studies of how the students use EULER and how this course affects their learning have enabled us to identify specific educational affordances and their potential application in supporting learning. This study describes thirteen educational affordances that were revealed through the responses given by students in an open-question survey and interviews as well as through transcripts from researcher observations and student discussions. These educational affordances of ubiquitous learning environment employed in the natural science course have implications for current and future developments in pedagogy.

Keywords

Collaborative learning, Educational affordances, Game-based learning, Natural science learning, Ubiquitous learning

Introduction

The term “affordance” was coined by the perceptual psychologist J. J. Gibson, who developed an “ecological” alternative to cognitive approaches (Gibson, 1977). Gibson (1977) defined affordances as what the environment offers to humans and what it provides or furnishes, which may be for good or ill; it emphasizes possible actions that the observer perceives as feasible in the environment. Norman (1988) defined affordances as opportunities for action, that is, the perceived and actual fundamental properties of a thing that determine its possible uses. Independent of perception, affordances exist whether or not the actor cares about them, perceives them, or has perceptual information about them (Gaver, 1996). These definitions suggest that the same environment perceived by different observers may produce different perceptions of affordances, which are determined by the culture, social setting, experience, and intentions of the observer (McGrenere & Ho, 2000). On the basis of the abovementioned definitions, affordances cut across the subjective/objective barrier. Affordances are objective in that their existence does not depend on value, meaning, or interpretation. Additionally, affordances are subjective in that an actor is needed as a frame of reference; the actor and the environment make an inseparable pair (McGrenere & Ho, 2000).

Kirschner (2002) further defined educational affordances as those characteristics of an artifact that determine if and how a particular learning behavior can possibly be enacted within a given context. Webb (2005) illustrated that the affordances of information and communication technologies (ICT) provide the following pedagogical effects in science learning, which are promoting cognitive acceleration, enabling a wider range of experience in science learning, increasing self-management of students, and facilitating data collection and presentation. Therefore, the educational affordances are worthy of being explored and discussed.

Related works and our work

Mobile learning that employs mobile devices as learning tools has been thriving since recent years. Many studies are exploring the educational affordances that mobile devices can offer. For example, Klopfer, Squire, and Jenkins (2002)

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.
proposed five educational affordances of PDAs: portability, social interactivity, context sensitivity, connectivity, and individuality. Patten, Sañches, and Tangney (2006) explored current mobile applications and used the functionality to describe what might be understood as the following seven educational affordances: administration, referential, interactive, micro world, data collection, location awareness, and collaboration. Churchill and Churchill (2008) did an empirical study and found PDAs providing five educational affordances: a multimedia-access tool, connectivity tool, capture tool, representational tool, and analytical tool. Churchill and Churchill (2008) claimed that understanding of educational affordances is important in the context of planning a suitable intervention to support pedagogically effective integration of PDA technology.

In this study, we define the learning package (LP) as the combination of computer-assisted learning environment and course that includes diverse learning tools, rich learning resources, theme-based learning activities, and real-life learning contexts. We argue that the educational affordances of an LP are obtained from four elements: (1) Learning tools: Diverse learning tools used in the computer-assisted learning environment can provide support for learning, thus helping learners to acquire knowledge. (2) Learning resources: Rich learning resources provided in courses can offer support for basic knowledge to learners, allowing them to integrate existing knowledge and access information to build new knowledge. (3) Theme-based activities: Interesting theme-based learning activities provide the basic learning scaffolds to learners by prompting issues to be studied, guiding them to finish tasks, providing ideas and the framework for conclusion, and developing their skills to solve problems. (4) Learning contexts: Real-life learning contexts can increase the usefulness of knowledge, induce learners’ motivation to learn, and reinforce learning experiences.

Our primary research question is as follows: What are the educational affordances of an LP? Although learners’ perceptions should be the best sources of educational affordances, this study uses a learning environment that includes diverse learning tools and rich learning resources as well as to conduct a natural science course that includes various theme-based learning activities in a real situation for exploring the educational affordances. This study also investigates the perceived degree of educational affordances that the students identified.

Methodology

A case study was conducted in a classroom and at Guandu Nature Park, a famous wetland in the Taipei area (Wild Bird Society of Taipei, 2009), for an eight-week natural science course. Qualitative and quantitative analyses were used to explore the educational affordances and perceived degree of educational affordances.

Course

The advantages of problem-based learning (PBL) include improvements in self-directed learning, self-motivation, problem-solving skills, and knowledge-application skills (Stepien and Gallagher, 1993). Game-based learning (GBL) has the potential to make learning more learner centered, easier, enjoyable, interesting, efficient, and effective than non-game-based learning does (Prensky, 2001; Virvou, Katsionis, & Manos, 2005). GBL can improve students’ learning achievement, learning motivation, attention, and creativity (Eow et al., 2009). Moreover, collaborative learning (CL) can increase learning motivation (Ushioda, 1996), improve achievements (Duin, 1984), and achieve group goals (Nichols, & Miller, 1994). Therefore, in order to discover educational affordances, this study develops a natural science course based on the PBL, CL, and GBL strategies that included several theme-based learning activities.

The course was entitled “How can the gradually disappearing wetland ecosystems in Taiwan be protected and restored?” The learning objectives of this course were as follows: (1) understanding wetland habitats and wildlife, (2) understanding the wetland ecological systems, (3) understanding the relationship between wildlife and environment, and (4) finding the methods of environmental protection. The course includes nine topics: (1) Functions and value of wetland, (2) Migrant, (3) Plant, (4) Crab, (5) Insect, (6) Mammalia, (7) Amphibia, (8) Fish, and (9) Invertebrate. The course includes the following five activities:

Activity 1. Understanding the ecology of wetlands (weeks 1–4)
Students perused the course material, browsed media resources, collected information, and shared knowledge to understand wetland ecosystems by using PDAs in the classroom.

Activity 2. Treasure hunt game (week 5)
Each student used PDAs with radio frequency identification (RFID) readers when visiting five scenic areas (riverside biological area, pond biological area, coastal forest area, crab watching area, and bird watching area) following the instructions on a treasure map.

Activity 3. Wetland rescue (week 6)
Each team selected a collaborative team project from among the predefined themes: “What is the interdependent relationship among biological species in a wetland?” “Why do wetlands gradually disappear?” “How to protect wetlands?” and “How to restore wetlands?” The team members were required to develop hypotheses, find evidence, synthesize knowledge, and explain phenomena outside the classroom.

Activity 4. Unexpected encounter (week 7)
Each student adopted mobile augmented reality (m-AR) to observe virtual creatures rarely seen in the wetland areas during the season.

Activity 5. Survey and interviews (week 8)
The teachers conducted an open-question survey with the students, and the researchers conducted in-depth interviews with the students.

Learning environment

The rapid development of modern computer science and sensor technology has enabled many new ubiquitous computing applications. Ubiquitous learning has the following primary characteristics: permanency, accessibility, immediacy, interactivity, situation, calmness, adaptability, seamlessness, and immersion (Liu, 2009). Many investigations employed the ubiquitous learning concept to enhance natural science learning (Tan, Liu, & Chang, 2007), share knowledge, provide help (El-Bishouty, Ogata, & Yano, 2007), and cultivate student problem-solving ability (Liu, Tan, & Chu, 2009).

This study employed a ubiquitous learning environment, EULER, to help natural science learning. EULER consists of two subsystems, Mobile Interactive Learning Environment (MOBILE) server, which is built on a laptop computer for teachers, and mobile-tools (m-Tools), which are built on PDA platform for students. Teachers input material into the Mobile Content Database (MCDB) via Mobile Content Management (MCM) and establish assessments in the Mobile Assessment Database (MADB) via Mobile Assessment Management (MAM). The personal portfolio tool, m-Portfolio, saves each student’s personal learning portfolio into the Mobile Learning Record Database (MLRDB) of the mobile server, making it available for teachers to review and assign grades. The MLRDB stores the student’s
records, including assignment grades, reading times, number of discussions, and instances of information sharing. Teachers preinstall the tests into the mobile server and conduct these tests using a MAM unit. Teachers use a Learning Activity Management (LAM) unit to construct a virtual classroom for each course, which supports many learning functions including bulletin boards, forums, voting, chatting, assignments, tests, and instruction. The m-Tools has fourteen functions, such as m-Messenger (receiving teacher guidance), m-RFID (processing RFID internal code), m-Loader (accessing course materials), m-Browser (browsing the contents), m-Player (playing multimedia files), m-AR (displaying virtual objects), m-Capture (recording videos and static photos), m-Notes (compiling information into articles), m-Share (sending data to peers), m-Test (taking tests), m-Calendar (organizing personal schedules), m-Voting (presenting personal idea), and m-Portfolio (recording personal learning records). The personal learning portfolio was saved in the m-DB. Figure 1 illustrates the system architecture of the proposed EULER (Liu, Tan, & Chu, 2009).

Participants

Participants included four natural science teachers and 36 fifth-grade students. All participant teachers had at least ten years of experience in computer-assisted instruction. Students begin to learn natural science from the first grade and Computer Science from the third grade. Therefore, students acquire basic skills in using information technologies to assist learning.

Procedures

Teachers demonstrated how to use the EULER functions and introduced the course activities to the students before the activities began. Figure 2 is a diagram briefly outlining the experimental procedures and data analysis.

In activity 1, the “Understanding the ecology of wetlands,” was undertaken in the classroom. By using their PDAs, the students perused course materials and browsed media resources from EULER, collected information from the Internet, and shared knowledge with their classmates in order to understand the creatures and environment of the wetlands.

In activity 2, “Treasure hunt game” activity, the teachers prepared numerous information boards, each with an RFID tag attached. The teachers established the relationship between learning materials and the identification codes of RFID tags and then placed the information boards near the corresponding wetland creatures. Each student carried a mobile device equipped with a video camera and an RFID reader when visiting scenic areas using the treasure map. Each student approaching a scenic area could use the learning device to identify the RFID tag. The detected identification code of the RFID tag was then transmitted via WLAN to the MOBILE server. The MOBILE server recognized the location of the student and then transmitted the context-aware content to the learning device. After completing a learning unit at one scenic area, the student answered four questions and was given a virtual gold coin; he/she then continued to the next scenic area until all five areas had been visited. The student thus accessed context-aware content related to actual wildlife, which enabled game-based ubiquitous learning. Figure 3 illustrates the scenario of treasure hunt game (Liu, Tan, & Chu, 2009).

In activity 3, “Wetland rescue” activity, the students selected a team project from among the themes. In case the students were unable to propose any feasible method even after group discussions, they used the messenger tool (m-Messenger) to request the teacher’s assistance. Under the guidance of the teacher, team members applied m-Capture to capture images or videos as evidential materials. Each team member organized and collected evidential materials and retrieved relevant information from the MOBILE server and then compiled these materials into an article. The members engaged themselves in a group discussion on the basis of the articles and drew an inference for a selected
problem. All articles were then sent to team leaders to edit a rich team report that was submitted to teachers by using sharing tool (m-Sharer).

Figure 3. The scenario of treasure hunt game

In activity 4, “Unexpected encounter” activity, each student utilized the AR tool (m-AR) to watch virtual wildlife that would rarely appear in the wetland area during the season. When students approaching a scenic area, the MOBILE server recognized the location of the student and season and then transmitted a virtual 3D animation of wildlife to the students’ learning device. Therefore, the students could understand wildlife that usually appeared in this area. For instance, students could see virtual mandarin ducks swim on a stream shown on the PDA screen through superposed virtual imageries on the live video.

In activity 5, on completion of the course, teachers and researchers conducted an open-question survey, a yes/no-question survey, and in-depth interviews.

Data sources

Data sources for this case included the students’ written responses to the survey questions, transcripts of in-depth interviews, on-line student discussions, and the researchers’ written observation notes. After the learning activities, a survey was conducted to explore the educational affordances of EULER. The survey included the following open-ended questions. (Q1) How did you utilize m-Tools to assist you in the “Understanding the ecology of wetlands” activity? (Q2) How did you utilize m-Tools to assist you in the “Treasure hunt game”? (Q3) How did you utilize m-Tools to complete the “Wetland rescue” activity? (Q4) How did you utilize m-Tools to complete the “Unexpected encounter” activity? Each student was required to respond to the aforementioned open-ended questions. The transcripts were checked by the students.

In-depth interviews were conducted to explore the students’ points of view, feelings, and perspectives after they completed all the learning activities. The researcher interviewed the students for 20 minutes to understand their opinions about the learning activities. The interviews comprised the following questions: (I1) Which activity do you think is the most interesting? Why? (I2) Which tool do you think is the most useful for aiding in learning? Why? (I3) From which activity do you think you gained the most? Why? (I4) What is your biggest gain from this course? Why? (I5) What new uses for m-Tools in the learning activities did you discover? The in-depth interviews were recorded on an audio recorder.

The data collected from on-line discussions were related to the students’ reflections on their uses and exploration of m-Tools. The researchers observed the students’ learning behavior and their use of m-Tools during the learning
activities, and they documented the learning processes and important findings in the form of notes over a period of eight weeks. The data was organized, and it classified the educational affordances relating to the proposed ubiquitous learning environment and course. Finally, we obtained 36 open-ended questionnaires, 28 interview records, records on 36 topics of discussion, 15 observation records, 6 team reports, and 36 yes/no questionnaires (response rate is 100%). Teachers and tutors helped students to understand the meaning of each question. The 36 response questionnaires comprised 218 responses and were collected from the students.

Data analyses

The data were collected, organized, and reviewed by researchers who manually use the text mining technique (Feldman & Hirsh, 1997) on features extraction and based on the following steps to obtain the educational affordances:

Step1. Each student’s response was assigned a code; for example, r-q3-s7 indicates the response to question#3 by student number 7.

Step2. Extract keywords about educational affordances manually from each student’s responses, such as the keyword “data collection.”

Step3. Arrange the distinct keywords that were regarded as educational affordances, from all extracted keywords.

In order to obtain the perceived degree of each educational affordance, in quantitative data analyses, we designed the yes/no questionnaire that asked students whether they used the tools/features/methods described in the questionnaire to help in learning activities. The perceived degree of educational affordance is defined as the number of students who reply “yes” divided by the number of students on an educational affordance.

Results and discussions

Analysis of data sources revealed eight actual educational affordances and five perceived educational affordances. We employed a representative student’s response that was paraphrased to enhance readability as a sample for each affordance, and the other students’ responses, interview records, and observation records were described in discussion paragraphs.

Actual educational affordances

Unconstrained Knowledge Accession—knowledge construction is the ability to recall specific and isolatable information from learned material (Bloom, 1956). M-Tools can enable students to retrieve learning materials from the content server at any time and place, students through reading, comprehending, and memorizing those materials to construct their knowledge. A student wrote the following in his survey response:

   Student 6: I saw many crabs move sideways. I wanted to know whether all crabs move sideways. I employed the browsing tool (m-Browser) to find the answer from MOBILE server, and I found that one of the species, the soldier crab (Mictyris brevidactylus), is an exception. Using these ubiquitous learning tools, I could easily acquire new knowledge.

In the researchers’ observation, the students were able to use the browser (m-Browser) and player (m-Player) to read materials and play videos. These materials helped them further understand the features, living habits, and feeding habits of animals.

Real-time Evaluation—evaluation is a treatment to assess the learning outcomes for achieving specific educational purposes (Anderson & Krathwohl, 2001). The test tool (m-Test) could enable students to take real-time tests and to improve their cognition. A student noted the following in his survey response:

   Student 21: Answering questions immediately on the PDA is a new experience. On an electronic test tool (m-Test), colorful graphics and sounds can be attached to each question. Compared with paper-and-pencil
tests, electronic tests are more lively and interesting.

During the interviews, the students stated that they thought using the m-Test to take tests was interesting, and the immediate retrieval of answers was helpful for learning. Therefore, real-time tests improve student outcomes and stimulate interest in learning (Liu, Tan, & Chu, 2009).

**Individuality**—individuality indicates separateness; individuality is the state or quality of being an individual, a student separate from other students and possessing his or her own educational needs, goals, and desires. Klopfer, Squire, and Jenkins (2002) argued that handheld computers provide a unique scaffolding function that is customized to individual paths of investigation. A student noted the following in his survey response:

> Student 31: During the learning activities, the portfolio tool (m-Portfolio) automatically recorded the materials I read, stored the graphics of crabs I captured, saved my homework, recorded the treasure hunt games I attended, and kept track of my test performance.

In the interview, a student stated that she recorded the deadline for each assignment on the calendar (m-Calendar). The calendar would automatically alert her one day before each deadline. Therefore, handhelds enable students to configure their own personal settings, profiles, activities, and private portfolios. The learning habits of students who habitually set goals, priorities, and well-organized learning schedules are positively related to average academic performance. Each student read different materials, watched dissimilar wildlife, stirred diverse reception, and procured a diverse learning experience that could be saved in portfolios.

**Diverse Interaction**—there are four interactions in ubiquitous learning: learner-content interaction, learner-instructor interaction, learner-learner interaction, and learner-environment interaction. Interaction in teaching and learning has a positive meaning for learning efficiency and developing thinking ability (Vygotsky, 1978). A student wrote the following in his survey response:

> Student 16: Whenever we encountered problems during the “Wetland Rescue” activity, we used the communication tool (m-Messenger) to chat with our teacher and ask for assistance. The teacher also used it to guide us and give hints to the solution. All problems were finally solved after group discussion.

On the discussion forum, a student stated that he posted the following question: “Why do crabs walk sideways?” This issue drew much attention and discussion. Many classmates provided their opinions. Some were interesting and creative. The students browsed classmates’ articles on the forum and shared their opinions. By exchanging opinions, they made many new friends. Therefore, through social interaction processes (e.g., exchange ideas and discuss issues), students improve their comprehension.

**Arbitrary Data Collection**—data could be collected from the students’ observation, recording, investigation, and measurement, or from existing videos, audios, images, and articles that were stored in the content management server. The learning environment enables students to collect data freely and helps cultivate their abilities. A student wrote the following in his survey response:

> Student 15: Whenever I observed anything interesting, I would easily use the video-capturing tool (m-Capture) to record the scenes in the PDA. For instance, I saw two crabs that seemed to be fighting, so I recorded them in a video file and uploaded the file to the MOBILE server, which made the video available for others to view.

In the discussion forum, a student said, “In the ‘Wetland rescue’ activity, I easily used m-Capture to take snapshots of the special things I observed, such as a crab digging up its cave, a crab swallowing sand, and a dragonfly perching on a leaf.” Therefore, as learning occurs in a rich natural environment, filled with real creatures, the concepts presented can be meaningfully contextualized.
Ubiquitous Game Play—ubiquitous games are developed using ubiquitous technology and game science; in the real environment, players can use devices or equipment at any time and location to play interactive games involving a portion of physical objects and a portion of virtual ones, which allow them to feel personally and physically involved in the games (Liu & Chu, 2010). Games could enable students to increase their learning interest by motivating students at any time and place (Prensky, 2001). A student noted the following in his survey response:

Student 18: In the “Treasure Hunt” activity, when we arrived at the crab watching area, we used the m-RFID tool to get crab-related course materials. We would receive a virtual treasure after we completed the learning and passed a test in each zone. Finally, our team received an award after collecting the five treasures in the shortest time.

In the researcher observation, the students really enjoyed the treasure hunt game, so each team tried to do its best to win the award. One student said, “Although we got virtual treasures in the treasure hunt game, we gained much priceless knowledge about environmental protection in the real world.” Therefore, the benefits of the game were increased motivation, enhanced collaborative learning activities, improved outcomes, maximized immersion, competition opportunities, and increased enjoyment (Liu, Tan, & Chu, 2009).

Authentic Context Awareness—contextual information can be used to characterize the situation of an entity, which can be a person, place, or physical object that is considered relevant to the interaction between a user and an application (Dey, 2001). The advantage of context-aware learning is to engage real learning experiences and improve learning effectiveness, especially in the real world (Cooper, 1993). A student wrote the following in his survey response:

Student 22: In the activity, I saw fiddler crabs waving fiddle claws and waterfowls looking for food beside the pond. I could also reach out to touch Kandelia candel. This activity not only allowed me to read webpage-based materials any time but also to watch biological creatures closely.

In the interview, a student stated that he was seldom exposed to nature. He thought that the learning activity allowed him to watch wild creatures and deepen his understanding. Therefore, context-aware learning enables students to gain authentic learning experiences and improve their cognition.

Vivid Immersion—immersion provides learners opportunities to learn by interacting with virtual objects, which, depending on the content, may enhance conceptual understanding of the content (Bricken & Byrne, 1993). A student wrote the following in his survey response:

Student 26: During the “Unexpected Encounter” activity, when I walked into the crab area, I could not see any Uca lactea (a species of fiddler crab). Through the m-AR function of the PDA, I could see a virtual Uca lacteal waving its fiddle claws on the swamp that are not normally seen in the real world and even those that no longer inhabit the earth.

In the interview, a student said, “With the augmented reality tool (m-AR), I could see a virtual dragonfly perching on a stone.” On the discussion forum, a student said, “Although I could not watch even a single Anas Platyrhynchos on the river, I could see a virtual one on the PDA screen by using the m-AR tool, which was very cool.” Therefore, immersive learning activity provides learners the opportunities to learn by watching virtual objects, which may enhance their conceptual understanding.

Perceived educational affordances

Skillful Application—application is the use of new knowledge to solve problems in new situations by applying acquired knowledge, facts, techniques, and rules in different ways (Anderson & Krathwohl, 2001). The perceived educational affordance of application is to provide learning tools to develop student thinking skills, to analyze questions, to solve problems, to work cooperatively, and to share personal experience. The students noted the following in their survey responses:
Students 10 and 11: My classmate Eddy and I saw thousands of crabs marching on the sand. He asked me “Do you know approximately how many crabs are there?” I immediately counted the number of crabs within a square of 20 centimeters (twelve crabs) and estimated the area of this sand land (about 12 square meters). Later, I used the calculator on the PDA to determine that there were about 3600 crabs on this sand land.

In the interviews, the students said, “After completing our report and submitting it to our teacher, our teacher immediately graded our report and presented it to other classmates.” Therefore, the students could apply the skills for sharing information, discussing issues, completing common tasks, and cultivating problem-solving ability.

**Methodical Analysis**—analysis means to break material into its constituent parts and to determine how the parts relate to one another and to an overall structure or purpose (Anderson & Krathwohl, 2001). Methodical analysis helps students enhance their problem solving skills. A student noted the following in his survey response:

Student 8: We could not observe how bacteria decompose fallen leaves and how planktons eat organic substances. On the basis of our data and observations, we concluded at the end of our discussion that the interdependent relationship in the wetland should be as follows: The fallen leaves will first be decomposed by the bacteria in the water or mud and then absorbed by planktons, which mollusks, shellfish, shrimp, crabs, and fish feed on. Mollusks, shellfish, shrimp, crabs, and fish constitute the diet of various bird species.

According to the observations, students could learn how to analyze a problem and then form a conclusion. Therefore, the analytical process emphasizes students’ abilities to differentiate, organize, and attribute abilities.

**Creative Synthesis**—synthesis is combining elements and parts to form a whole. This involves the process of working with pieces, parts, elements, etc. and arranging and combining them in such a manner as to constitute a pattern or structure which was unclear or not explicit before (Bloom, 1956). Creation is to combine elements to form a coherent or functional whole, compiling information in different ways by combining elements in new patterns or proposing alternative solutions (Anderson & Krathwohl, 2001). A student noted the following in his survey response:

Student 32: The test results of our team were unsatisfactory. However, we were good at using the image capturing tool (m-Capture), audio recording tool, and editing tool (m-Notes) provided by the system, so our report was highly rated. With this report, our overall performance was not too much behind that of others.

In the interview, students stated that they utilized the editing tool (m-Notes) to integrate collected files into a vivid multimedia-based report comprising text, sound, graphics, and video. Therefore, the editing tool helps the skill of student to synthesize.

**Ubiquitous revision**—if every student is given sufficient time to learn and is instructed with proper teaching tactics, all students can achieve the same learning achievement (Bloom, 1968; Levine, 1985). Ubiquitous learning environment provides students many more opportunities to review course materials. A student noted the following in his survey response:

Student 7: I stored class handouts and teaching materials in the PDA and reviewed them while waiting for buses. These tools allowed me to review lessons at any time and any place, and they deepened my understanding and memory of the lessons.

In research observation, handheld learning devices offer advantages in portability, storage capability, and multimedia functions. Students usually reviewed learning materials stored in PDAs during leisure hours, thus increasing their familiarity with the lessons.

**Seamless Collaboration**—collaborative learning means that students learn together, share experiences, solve
common problems, and complete assigned tasks (Zheng, Ogata, & Yano, 2004). Through diverse learning tools in the ubiquitous learning environment, the students could synchronously and continuously complete a common task in a distant place. A student noted the following in his survey responses:

Student 7: The title of our team report was “How to restore wetlands?” Under the guidance of the teacher, Melody captured the image of bank-protecting mud slopes and government beautification efforts. Carol found on the Internet that the government is currently dismantling unused fish ponds, filling them, and then removing waste on the land so that native plants can grow on the wetlands and provide a living environment for animals. Becky reported that the education center in the Guandu Nature Park was established to promote environmental awareness and is effective for teaching citizens to preserve the ecological system. We concluded that promoting the concept of environmental protection and artificially restoring the natural environment are the most effective methods of restoring wetlands.

On the basis of the abovementioned data analysis, we conclude that our proposed ubiquitous learning environment and natural science course affords eight actual educational affordances: unconstrained knowledge accession, real-time evaluation, individuality, diverse interaction, arbitrary data collection, ubiquitous game play, authentic context-awareness, and vivid immersion. The students identified five perceived educational affordances: skillful application, methodical analysis, creative synthesis, ubiquitous revision, and seamless collaboration.

The perceived degree of educational affordances

In the students’ perceived degree of educational affordances, the statistical results of the questionnaire is presented in Table 2 and illustrated in Figure 4.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes Degree</th>
<th>Educational Affordances*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do the learning tools help you to understand wetland ecosystems and environmental protection?</td>
<td>81%</td>
<td>Unconstrained Knowledge Accession</td>
</tr>
<tr>
<td>Do the learning tools enable you to participate in tests after finishing learning in a zone?</td>
<td>100%</td>
<td>Real-time Evaluation</td>
</tr>
<tr>
<td>Do the learning tools help you record personal learning portfolio?</td>
<td>75%</td>
<td>Individuality</td>
</tr>
<tr>
<td>Do the learning tools and course help you interact with teachers and classmates?</td>
<td>86%</td>
<td>Diverse Interaction</td>
</tr>
<tr>
<td>Do the learning tools enable you to record videos and photos anytime?</td>
<td>100%</td>
<td>Arbitrary Data collection</td>
</tr>
<tr>
<td>Do the learning environment and course enable you to play a game outdoors?</td>
<td>69%</td>
<td>Ubiquitous Game Play</td>
</tr>
<tr>
<td>Do the learning environment and course help you gain useful knowledge in a real context?</td>
<td>83%</td>
<td>Authentic Context-awareness</td>
</tr>
<tr>
<td>Do the learning tools enable you to watch virtual creatures in real world?</td>
<td>61%</td>
<td>Vivid Immersion</td>
</tr>
<tr>
<td>Do the learning tools help you solve the assigned problem?</td>
<td>89%</td>
<td>Skillful Application</td>
</tr>
<tr>
<td>Did you use the calculator in this course?</td>
<td>25%</td>
<td>Methodical Analysis</td>
</tr>
<tr>
<td>Does the learning environment help you finish the report and homework?</td>
<td>72%</td>
<td>Creative Synthesis</td>
</tr>
<tr>
<td>Do the learning tools help you review the course materials at any time and place?</td>
<td>58%</td>
<td>Ubiquitous Revision</td>
</tr>
<tr>
<td>Do the learning tools help your team finish team report?</td>
<td>83%</td>
<td>Seamless Collaboration</td>
</tr>
</tbody>
</table>

*This field was concealed in questionnaire.

The statistical results reveal that the real-time evaluation was approved by all students (100%) because they concentrated on participating in tests to evaluate their outcomes. The educational affordance on arbitrary data
collection was also approved by all students (100%) because they employed the EULER to collect data in order to obtain good grades on their homework and reports. Those two results prove Webb’s (2005) argument: ICT provides pedagogical effects in promoting cognitive acceleration and facilitating data collection. The educational affordances on unconstrained knowledge accession, diverse interaction, authentic context-awareness, skillful application, and seamless collaboration were approved by more than 80% of the students because most students used EULER to acquire knowledge, exchange files with members, share their experiences with classmates, interact with teachers, observe real wildlife, and solve the assigned problem. The educational affordances on individuality and creative synthesis were also approved by 70%–80% of the students; those who did not approve did not use EULER to organize their personal schedules and felt that their products were not creative. The educational affordances on knowledge construction, interaction, context-awareness, collaboration, and individuality also prove the finding from Klopfer et al. (2002), Patten et al. (2006), and Churchill and Churchill (2008). The educational affordances on ubiquitous game play and vivid immersion were approved by only 60%–70% of the students because the other students felt that the course designer could add more interesting games to the course and that the PDA screen was small and unclear when viewed in direct sunlight. The educational affordance on ubiquitous revision was approved by more than half the students (58%); those who did not approve were used to reading textbooks when reviewing. The educational affordance on the methodical analysis was approved by only a few students (25%); the reason was that this course did not design any assignment that required a methodical analysis to solve a problem. Several students raised mathematical or natural problems and then found the solution by using the learning tools. The educational affordance on analysis proved the finding of Churchill and Churchill (2008).

Figure 4. Bar chart representing the perceived degree of identified educational affordances

The organization of the aforementioned results and discussion and the relation between the identified educational affordances, learners’ actions, and educational goals are summarized in Table 3. Through the ubiquitous learning environment and natural science course offering the educational affordances, the learners’ actions enabled their learning to achieve the educational goals they pursued.

<table>
<thead>
<tr>
<th>Educational Affordances</th>
<th>Tools/Activities</th>
<th>Learner’s actions</th>
<th>Educational goals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Actual affordances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unconstrained Knowledge Accession</td>
<td>m-Loader, m-Browser, m-Capture, m-Notes, and m-Player/Understanding</td>
<td>Access materials, read lectures, compile articles, and play medias</td>
<td>Construct knowledge, build cognition, and enhance comprehension</td>
</tr>
</tbody>
</table>

Table 3. Summary of educational affordances of a ubiquitous learning environment in a natural science course
<table>
<thead>
<tr>
<th>Real-time Evaluation</th>
<th>m-Test, m-Notes, m-Voting, and m-Portfolio/ Treasure hunt game</th>
<th>Evaluate outcome, examine homework, examine reports, present intend, and evaluate portfolio</th>
<th>Evaluate learning outcomes, and improve knowledge cognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuality</td>
<td>m-Calendar, m-Portfolio, m-Browser and m-Player/ Understanding the ecology of wetlands</td>
<td>Organize personal schedule, save learning records, and manage personal portfolio</td>
<td>Perform individual activity, develop private portfolio, engage individualized learning experiences, accomplish learner-centered learning, foster self-management ability, foster good learning habits, and improve learning performance</td>
</tr>
<tr>
<td>Diverse Interaction</td>
<td>All tools/ All activities</td>
<td>Discuss issues, share information, present ideas, and present intent, and interact with virtual objects</td>
<td>Enhance peer interaction, increase motivation, improve outcome, improve human relationship, healthful psychology, improve learning efficiency, improve teaching quality, and develop thinking ability</td>
</tr>
<tr>
<td>Diverse Interaction</td>
<td>All tools/ All activities</td>
<td>Discuss issues, share information, present ideas, and present intent, and interact with virtual objects</td>
<td>Enhance peer interaction, increase motivation, improve outcome, improve human relationship, healthful psychology, improve learning efficiency, improve teaching quality, and develop thinking ability</td>
</tr>
<tr>
<td>Diverse Interaction</td>
<td>All tools/ All activities</td>
<td>Discuss issues, share information, present ideas, and present intent, and interact with virtual objects</td>
<td>Enhance peer interaction, increase motivation, improve outcome, improve human relationship, healthful psychology, improve learning efficiency, improve teaching quality, and develop thinking ability</td>
</tr>
<tr>
<td>Ubiquitous Game play</td>
<td>m-Browser, m-RFID, and m-Test/ Treasure hunt game</td>
<td>Retrieve context-aware materials, and play ubiquitous game</td>
<td>Increase learning interest, engage learning motivation, enhance learning interactivity, and improve learning outcomes</td>
</tr>
<tr>
<td>Authentic Context-awareness</td>
<td>m-RFID and m-Player/ Treasure hunt game</td>
<td>Retrieve context-aware data, play media, combine courses with real world, and observe wildlife behavior</td>
<td>Engage learning experiences, improve learning effectiveness, and construct authentic knowledge</td>
</tr>
<tr>
<td>Vivid Immersion</td>
<td>m-AR/ Unexpected encounter</td>
<td>Combine virtual and real world objects as well as provide spatial, temporal and contextual conceptualization</td>
<td>Improve understanding and enhance learning experiences</td>
</tr>
</tbody>
</table>

**2. Perceived affordances**

<table>
<thead>
<tr>
<th>Skillful Application</th>
<th>All tools/All activities</th>
<th>Access materials, browse materials, collect data, compile articles, play media, calculation, analysis data, present intend, organize schedule, manage personal information, discuss issues, share information, evaluate achievement, present ideas, retrieve context-aware materials, and compile reports</th>
<th>Apply information technologies for learning and foster problem-solving skill, exploring ability, and independent thinking ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodical Analysis</td>
<td>Embedded calculator and analysis tool/ Wetland rescue</td>
<td>Collect data, observe wildlife, analyze data, and make a conclusion</td>
<td>Foster data analysis ability and explore new knowledge</td>
</tr>
<tr>
<td>Creative Synthesis</td>
<td>m-Loader, m-Capture and m-Notes/ Wetland rescue</td>
<td>Integrate learning objects into productions</td>
<td>Combine elements and produce creative products, and achieve self-satisfactory</td>
</tr>
</tbody>
</table>
## Conclusions and future work

This study finds thirteen educational affordances divided into two categories. The ubiquitous learning environment and course provided eight actual educational affordances, which included unconstrained knowledge accession, real-time evaluation, individuality, diverse interaction, arbitrary data collection, ubiquitous game play, authentic context-awareness, and vivid immersion. The students' responses identified five perceived educational affordances, which were skillful application, methodical analysis, creative synthesis, ubiquitous revision, and seamless collaboration. Among these affordances, seven gained good approval (greater than 80%), four gained general approval (greater than 60% but lower than 80%), and two gained no-good approval (lower than 60%) from the students, implying that the educational affordances provided by the EULER and the natural science course gained major approval (average is 76%) from the students.

The educational affordances explicated in this study should be useful to teachers when they employ ubiquitous computing technology, rich educational resources, and diverse pedagogical tactics to improve the student learning effect and help students achieve their learning goals. Our future research will be dedicated to investigating the differences in educational affordances between different instructional environments and open courses using different learning tools, learning contexts, learning affections, theme-based activities, and instructional strategies.

## Acknowledgments

This study was sponsored by the National Science Council of Taiwan. We are very pleased to thank the elementary school, Affiliated Experimental Elementary School of Taipei Municipal University of Education, whose students and teachers participated in this research.

## References


Contextualizing a MALL: Practice Design and Evaluation

Yaming Tai
Department of Children English Education, National Taipei University of Education, Taiwan // yaming@tea.ntue.edu.tw

(Submitted September 12, 2010; Revised March 9, 2011; Accepted May 10, 2011)

ABSTRACT
Mobile technologies have been successfully implemented in language learning, and have supported various innovative designs. However, many of these designs require considerable technical knowledge and support that are beyond the workload capacity of most language teachers. This paper discusses a study using an over-the-market mobile device combined with a task-based approach to design a contextualized mobile assisted language learning (MALL) practice. Meanwhile, a small-scale study was conducted to evaluate the effects of the proposed practice. The results provide evidence suggesting that contextualizing MALL practice can improve language learning and related attitudes for learners. The primary goal of this study is to demonstrate an affordable MALL design for language teachers. The study serves as a reference for teachers in developing this new type of MALL practice to motivate students and enrich language learning.

Keywords
MALL, Learning context, Task-based learning, EFL

Introduction
Many see mobile learning as the next generation of learning (Sharples, 2000). Mobile technologies applied to language learning include PDAs, multimedia cellular phones, MP3 players, and digital dictionaries (Zhao, 2005). The appeal and benefits of these mobile or handheld devices supporting mobile assisted language learning (MALL) appear to be their portability, the ability to play and record audio, and cost efficiency as compared to that of a laptop or desktop PC (Wishart, 2008). Some researchers claim that the portability and accessibility of mobile devices can allow language learners to access learning materials and to communicate with teachers and peers with less time and space constraints (Chinnery, 2006; Nah et al., 2008; Rosell-Aguilar, 2007).

As communication has always been a central pillar of language learning, the technical aspect of communicative mobile technology has become increasingly able to support communication pedagogy in recent years. Specifically, MALLs support language learning in different language areas such as vocabulary learning (Cavus & Ibrahim, 2009) as well as pronunciation practice (Ducate & Lomicka, 2009). Some researchers have also drawn attention to the four language skills; for example, writing ability (Morita, 2003), and English listening skills (Edirisingha et al., 2007). These studies have adopted the wireless delivery mechanism of SMS and handheld functionalities of video and audio recording, as well as the playback of mobile devices to help learners practice specific types of second language skills.

In general, these adopted applications, because of the mature development of wireless technology and ICT devices, do not require much technical support, aside from uses in business domains. Language learning in respect to MALL focuses on practicing specific elements of knowledge and skills rather than using language merely as a means of communication.

Another area of MALL studies has explored the advantages of mobile technologies. These features include personal, situated, authentic, spontaneous, informal, and continuous access, as well as unhindered interaction across diverse contexts (Kukulska-Hume, 2009). For example, previous studies have examined a mobile peer-assisted learning system for a collaborative early English as a Foreign Language (EFL) reading program (Lan et al., 2007), a personalized intelligent m-learning system for supporting effective English reading (Chen & Hsu, 2008), and several theme-based m-learning activities improving contextual language learning experiences (Tan & Liu, 2004). MALL designs in these studies have emphasized innovative learning features of mobile devices, but all of them must customize mobile technologies. Although these researchers have proposed advanced MALLs, their designs and implementations require too much technical knowledge and too many demands that are not affordable for most language teachers. To apply MALL to real pedagogic situations, the main purpose of this study is to explore how teachers can adopt mobile technologies without too much technical burden, while enhancing target language acquisition to motivate learners.
Because communication in a target language is crucial to second language acquisition, language teachers must provide learners purposeful contexts where they can engage in authentic interaction. Among various versions of communicative language teaching, task-based instruction has become a concrete realization for developing target language through meaningful communication (Littlewood, 2004). With the communicative applications of advanced mobile technologies, mobile devices thus provide a platform for various communicative tasks. This study employed mobile technologies to design task-based language instruction, whose design and implementation must be affordable to language teachers. The instructional design used mobile technologies to support MALL in technically simple ways associated with sophisticated pedagogies. In this study, technology used in a simple way means that the technology itself is well developed and technically supported by the business domain; therefore, its incorporation into MALL does not require much technical knowledge or support from a language teacher. As the design is explored from the perspective of pedagogy instead of the functions of technology, this study asserts that the technical features of mobile technologies be aligned with the characteristics of the subject domain, EFL. The primary goal of the study is to explore how MALL can be designed in a manner that is technically simple for teachers and pedagogically motivating for students.

**Literature review**

According to Pica (2008), task-based instruction involves “activities that engages language learners in meaningful, goal-oriented communication to solve problems, complete projects, and reach decisions” (p.71). Based on the theoretical underpinnings of task-based language teaching, various tasks have been designed and their effects have been explored (Shehadeh, 2005). Research on the interaction hypothesis focused on the effect of the negotiation of meaning (Mackey, 1999). Research regarding output hypothesis examined the variables which affect learners in producing language output (Swain, 2000). The cognitive perspective emphasizes on what aspects of task can promote language fluency, accuracy, or complexity in language learners (Ellis, 2000). The socio-cultural position looks at how learners’ individual differences affect their task performance (Shehadeh, 2005). While these studies have analyzed the variables of tasks and explored their effects systematically, the issue of creating a favorable context for authentic communicative tasks to occur has been seldom discussed. Although context has been recognized as an integral factor in language teaching, serious attempts to provide students with natural contexts that prompt communication have been lacking. Widdowson (1998) indicated that contextual conditions must be authenticated by the learners, which occasionally requires some type of pedagogic artifice. Mobile technology seems to be an aid for teachers to make the target tasks learners are supposed to perform closer towards reality.

Advanced mobile technologies provide a wealth of ideas and strategies to enhance the design and implementation of foreign language teaching and learning. One line of study has employed mobile technologies to deliver content necessary for enhancing language learning. These include supporting vocabulary learning (Lu, 2008), advancing reading comprehension ability (Chen & Hsu, 2008), enhancing writing ability (Morita, 2003), improving English listening (Nah et al., 2008), and practicing pronunciation (Godwin-Jones, 2008). However, language learning context has not been a major concern for these studies. Another line of research has elaborated the features of mobile technology regarding MALL design in a context-aware manner. For example, previous studies have proposed a personalized mobile English vocabulary learning system for recommending appropriate English vocabulary materials to learners (Chen & Chung, 2008), and a mobile intelligent tutoring system with learners’ location awareness supporting language learning (Cui & Bull, 2005). Context is an essential issue of these studies, but it has been discussed more from the adaptive perspective rather than the facilitative perspective. With regard of using task design with mobile technology, Kiernan and Aizawa (2004) evaluate the use of mobile phones as tools for university students English learning and suggest that mobile phones should be a language learning tool worthy of further investigation.

**Proposed design and related theories**

Based on second language acquisition theories, the goal of task-based approaches is to engage students in interactions to facilitate second language development. Therefore, the employment of mobile technologies to design tasks must emphasize the potential of mobile technology to promote social interaction. This study considered Long’s (2000) proposed steps for designing task-based language teaching. The first step is to identify target tasks based on learner needs and language proficiency. The target tasks were thereby classified into task types. Finally, pedagogic
tasks were developed from task types. The potential relevance of mobile technologies was identified to map the individual task. In other words, the characteristics of the pedagogic tasks were designed by ready-to-use mobile technologies to amplify the communicative aspects of the interactions. As shown in the design, mobile technologies were neither considered nor applied until the types of tasks had been decided.

This study proposes three major MALL designs that are technically affordable for a language teacher and facilitate authentic language communication. The first design involves creating task authenticity. The second design entails providing the co-existence of social and digital participation. The third design involves providing dynamics of communicative mediations. Simple technologies are employed to construct the tasks. For creating files as the instructional materials, teachers use popular software, such as Photoshop to edit photos as well as pictures, and Goldwave to create audio files. As the communicative tasks are enacted, teachers use the applications of software in the mobile device to engage the task, which include the internet access, calling, MMS, SMS text-messaging, and MSN.

With regard to task authenticity, Schank and Kass (1996) revealed that, when students try to accomplish task goals within authentic contexts, they tend to be highly motivated to learn. The focus of authenticity here, following the view of Barab et al. (2000), is not on the task itself, but on the real communication that occurs among learners. Authenticity lies in the dynamic interactions among the learner, the task, and the environment.

For creating task authenticity, inspired by the mixed-reality in mobile learning, mobile device and contextual objects around students are used together to provide task-related information for guiding and fostering students’ real communication. An authentic context is required for learners to perform an authentic task because anchoring the task in a meaningful and realistic context is vital. The context includes the location, identities people and objects nearby, and environment around those objects. For bridging the gap created by the digital information and reality, learners are stimulated to produce language output. This design should support the output hypothesis. That is, based on understanding of authentic artifacts, acquired digital information, and peer assistance as input, students are assisted to exercise language skills and ask for further clues while having collaborative discussions as output.

As for the technological design, instead of using sensoring technology such as radio frequency identification (RFID), this study is stimulated by Chinnery’s (2006) discussion about moblogging. Moblogging uses a cell phone or PDA in the field to post words or pictures to a website, and offers the potential to add authentic and personal visual content. This study proposes the opposite, using pre-scripted tasks for specific contexts in which students must rely on both mobile devices and the contexts to complete the tasks. While working on the task, the contexts and digital information in the mobile device form a mixed-reality or augmented reality MALL. The technical requirements are only the store and display of multimedia information in the mobile devices, as well as their communicative functional operations for voice call and messaging; all of which should be technically affordable for most language teachers.

For providing the co-existence of social and digital participation, mobile devices contain two distinct types of participation, which occur simultaneously in the same place. When students are engaged in a task, they can experience normal social participation within their group. They can also send and receive digital information from other people in or outside of their group. These two types of participation regarding the mobile learning features of social co-existence and digital participation can stimulate students’ emotional expressions and corresponding social language.

The proposed design requires learners to acquire the instructions for their task through the digital information. While they are collaborating with each other to achieve this goal, they can use the mobile device to inquire about or provide information. After gathering information for their assigned task, the group members must exchange and synthesize the information to work out the task. As a group, they must not only use target language to communicate with each other for collaboratively finishing the task, but must also gather any necessary and specific information, such as feedback or help from their instructors. As collaboration is viewed as elemental to task-based instruction, students may encounter the challenges of role distribution, low motivation or even futility. Instructors can perform monitoring or scaffolding as needed to keep the momentum of the task.

This type of co-existence of participation is somewhat distinct from that of science learning, for example, the spread of disease (Zurita & Nussbaum, 2004), in which students’ social interaction is used to represent the science
curriculum to be learned. In language learning, the interaction hypothesis emphasizes the interaction between learners for acquiring a second language (Mackey, 1999). The co-existence of social participation and information serve as a platform on which students can practice social-linguistic expressions and create the needs of interaction between learners and other speakers.

For providing dynamics of communicative mediations, the mobile device is used as a mediator to facilitate output and promote interaction among students and teachers. The common features of mobile devices include internet access, voice-messaging, SMS text-messaging, and MSN. Hence, the communication can be in spoken or written forms, which may require learners to use their reading, writing, listening, or speaking skills according to the nature of the task. In addition, mediated communication also allows both asynchronous and synchronous communication. These forms of communication can be viewed as task characteristics from the cognitive approach (Ellis, 2000) and can affect learner performance.

The aforementioned designs stress that English is a language for communication, not just an object of academic interest or merely a key to passing examinations. The proposed design integrates MALL into task-based language learning by emphasizing the communicative nature of mobile devices and their capability for contextual use facilitated by their portability and accessibility.

**A sample practice design**

**The mobile device**

The mobile device chosen for this study has dimensions of 120x70x20 mm and five features advocated by the vendor: mobile MSN with full keyboard input, POP3 email, Internet browsing and RSS news feeds, MP3 player, and basic mobile phone functions of voice and SMS. These features form a comprehensive, communicative functionality, that is, listening, speaking, reading, and writing, supporting this study’s proposed MALL.

The operation of the mobile device to perform the task has been evaluated to be affordable for the learners. This is critical because the mobile device is viewed as an interaction platform for learners to handle both their language input and output, and it should not require too much effort to use. If a device requires too much demand of attention for its operation, less attention can be devoted to the learning it supports.

**Three phases of the task**

A three-phase task framework is proposed. The first and third phases are intended for the classroom as pre- and post-task learning; and the second phase is the main task. The aim of such design is to begin with the preparation for the task in the classroom, followed by communicative language practice in the field. The field visit is then followed by a reflection in the classroom with presentation and sharing of what the students have learned. Three phases form a combination of two modes of instructions (face-to-face and MALL) and help students experience language learning and usage in varied contexts. The main task is conducted at the Lin Family Mansion and Garden in Taiwan. This field was chosen because each building in the garden has particular features and routes that are like a maze. Additionally, the area is large enough so that teams will not interfere with one another when performing their tasks.

**Pre-task phase**

The pre-task involves introducing the task with three functions. These include motivating the learners to perform the task; preparing the learners to perform the task; and providing clear instructions on how the task should or can be performed (Gorp & Bogaert, 2006). The linguistic content required to perform the task is introduced for learners to recall or learn. The vocabulary and sentence structures are from the Nine-Year Integrated Curriculum published by the Ministry of Education in Taiwan. Both vocabulary and structures are selected based on their relevance to the task. In addition, ten new vocabulary words relevant to the task are also taught.

The mobile devices are used as classroom response systems to offer personal interactions with language learning
experiences in this phase. Moreover, because the mobile devices are going to be integral in scaffolding students’ task performance, participants are shown where they can find the referential materials in the mobile devices. The use of mobile devices in this phase entails students performing the simplified task and motivating them because these learning activities, such as using the mobile device to present learning material and communicate with the instructor one-on-one, are novel to them. The scenario regarding the task is presented in a film clip to show the students what they must do to complete the next task.

Regarding mobile device operation, because not every student is familiar with the use of mobile devices, introduction of and practice using the mobile device functions are necessary. Therefore, activities in this phase are designed by combining the linguistic content (input) and the functions of the mobile device. For instance, students learn to open a music file to listen to a song and use the mobile function of messaging to send answers to the instructor.

Main task phase

This phase is designed to generate authentic interaction, discussion, and negotiation among language learners. Three or four participants are formed in one group with a foreign agent. Students are assigned to various roles with different job contents. The jobs are distributed according to the mode of communication, a caller, an SMS reporter, an MSN reporter, and a secretary for recording and requesting help. Participants must cooperate with team members and must be able to understand the described task by listening to or reading the English instructions. They must then formulate responses in English either in spoken or written form. For instance, they must create their own sentences to describe the differences in a room compared to the picture file shown in the mobile device.

The task is designed as a problem-solving task. The learners are told that a burglary has been committed at the Lin Family Mansion and Garden. Because this burglary might be related to many other museum burglaries in other countries, an international organization is here in Taiwan to help the police find the suspect, but these agents only communicate in English. Some evidence has been collected by the police. The task for the students involves using this evidence to help the agents solve the crime and identify the suspect. Students must accomplish six task assignments. All evidence is designed according to particular context of different locations in the Garden. Participants are able to use mobile phones to access multimedia content linked to specific information for the task at different locations. As players report the accomplishment of a task, information about the introduction of a new task assignment is delivered to their devices for moving to different locations. They must use this information to progress in the game. Related picture files and sound files are pre-saved in the mobile device. The tasks require various language forms and mobile device skills and should be performed as a team, which can prompt learners to cooperate with each other to solve problems. After completing all six assignments, students must synthesize the results from each task and identify who the burglar is. For communicating with headquarters, the mobile device is used as a communicative tool as well as a supporting aid for teachers to send guidance and feedback to the learners. The six-task assignment, the sentence structures and language skills students might use are listed in Table 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Task Description</th>
<th>Mobile Technology</th>
<th>Language Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spot the differences</td>
<td>A picture of a portrait has been found. The police are wondering whether the suspect might have switched the portrait. Open the picture file with the title portrait. You need to locate the portrait and find the differences between the one you find and the one in the picture file.</td>
<td>Use MSN to report</td>
<td>Reading / Writing</td>
</tr>
<tr>
<td>Interview the witness</td>
<td>A foreigner was found in the pavilion near the pond. The police doubt she might be the witness in the crime scene. You need to interview her to get her personal information.</td>
<td>Call the headquarters to report</td>
<td>Listening / Speaking</td>
</tr>
<tr>
<td>Seek the source of sound</td>
<td>A janitor heard some strange sounds last night. Luckily, she used her cell phone to record the strange sound. Open the music file. You need to Call the headquarters to report.</td>
<td>Listening / Speaking</td>
<td></td>
</tr>
</tbody>
</table>
help the police to find out where the sound came from and any other clues in the recording.

<table>
<thead>
<tr>
<th>Locate the place</th>
<th>A map has been found. The police suspect that it shows the area where the burglar got into the Garden. Open the picture file with the title map and find where it is.</th>
<th>Use SMS to report.</th>
<th>Reading / Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decode</td>
<td>A scenic spot stamp with two numbers (4, 6) was found on the left arm of one suspect. Open the picture file with the title stamp. Find out the meaning of the codes.</td>
<td>Use SMS to report.</td>
<td>Reading / Writing</td>
</tr>
<tr>
<td>Check the items</td>
<td>Some items seemed to have been stolen. Go to check the objects in the room and describe the items to the police.</td>
<td>Use MSN to report.</td>
<td>Reading / Writing</td>
</tr>
</tbody>
</table>

The mobile device in this study is embedded with a variety of tasks, aimed at enhancing interactivity and engaging learners in using their English. Tasks are all centered on information gaps, problem-solving gaps, and reasoning gap, based on a burglary scenario. By employing the MALL design, the mixed reality is used to create the task. Considering the task “Spot the differences” as an example, the learners would see a digital photo of a portrait with several changes made on their mobile device and would compare it with the actual one in reality. As for the task “Decode” is regarded as complex for the learners, and the design of co-existence of social and digital participation is used. During the task, learners must interact with each other for discussion; meanwhile, the learners must keep connection with the person at headquarters using the mobile device to check if there are more clues for them to solve the problem. The dynamics of mediators are explored in the various tasks. Considering the task “Check the items” for example, the learners must report by MSN because they must describe the items in detail. In addition, the location is a study room; learners are told that MSN is the more effective way to communicate in this situation.

By actively solving the burglar case, learners need to use integrated language skills, collaborate with team members, interact with the instructor in headquarters. With the tasks embedded in the scenario, the goal of the task for students may be to find the real burglar, but the real purpose of the task is to engage learners in generating authentic communication. For instance, in the task ‘Seek the source of sound,’ students need to read a story of the janitor about what happened to her the night before and listen to a recording. According to the clues, students discuss with each other. As they find out the place, they need to report to headquarters by calling. Students need to integrate their reading, listening, speaking skills to accomplish the task.

**Post-task phase**

The post task helps students review and compare what they have learned in the pre-task and what they used during the task. The mobile device, hence, is integral in presenting the materials and connecting the two phases. Tasks are designed to create an environment in which learners are allowed to experiment with language, use language functionally, and make mistakes while doing so. A post-task aims at a verbal and interactive reconstruction of the process of task performance (Gorp & Bogaert, 2006). If learners encounter some communicative problems during the task, these problems create a “need to know” situation. Because learners' answers are delivered through the mobile device, their various answers have been recorded and are presented in this phase. The instructor in this phase collects as well as shows the various responses to the learners for them to compare and reflect.

**Evaluation**

**The instruments**

**Proficiency test**

A pre- and post-test is conducted to assess the learners’ English proficiency. The test consists of two sections, involving vocabulary and sentence structures. Twenty items are in each section. The test items are task-relevant vocabulary and sentence structures. A pre-test is designed and administered to assess learner proficiency. The same
test is administered again right after the activity to evaluate their learning performance.

**Attitude toward the design**

Because learners’ attitudes regarding their learning situations greatly affect their learning achievement (Gardner, 2000; Nah et al., 2008), this study evaluated how participants reacted to or felt about this design. In addition, this evaluation was intended to be comprehensive, because it is the first type of such learning in the field of MALL.

The pre-activity surveys included students’ attitudes toward English learning in general and English learning through mobile devices. A post-activity survey investigated learners’ attitudes toward the proposed learning activity. Their responses were evaluated based on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”.

**Participants**

The 35 participants, 21 boys and 14 girls, were sixth graders at the time of this study and had been learning English for three years with 80 minutes of structured English per week. Participants were grouped into 11 teams, that is, each team contained three or four members. The grouping was heterogeneous based according to the pre-test scores of the learners. None of these children had experience using mobile phones to learn English. However, 10 of the 35 students had experience using computers to learn English.

**Results**

**Observation**

The observations data included two snap shots of team activities. The purpose was to provide direct evidence about how participants performed their tasks, and to triangulate the quantitative data. In Figure 1, the female student is calling the help center to obtain further assistance regarding her group’s task in that location. Meanwhile, the other two students are interviewing the woman (played by an instructor) with white hair, seated on the left. The task the participants were trying to accomplish involved asking the lady about the burglary, particularly, inquiring about any suspicious people she might have seen the previous night. One of the participants is calling headquarters to obtain required questions and reiterate the questions to her team members. They then needed to call headquarters and give them the answers. The mobile device is used as a tool for communication. Participants could not have face-to-face communication with the instructor; therefore, authentic English output needed to be produced.

![Figure 1. Interviewing the witness](image)

The participants in Figure 2 are reading a message displayed on the mobile device, trying to figure out the clues provided for that specific location. Meanwhile, one of the participants is pointing to the text while talking with the other teammates. Viewing language learning as a social practice, and it is assumed that language learning is most effectively served by collaboration. Collaboration among participants in this study involved identifying some linguistic problems and performing a task. Moreover, through their collaboration, participants produced output to report to headquarters. The pedagogical use of this mobile device capability included the co-existence of social and
digital participation. Through participants’ collaboration, they could reveal the content of learning and construct knowledge in the realm of social experience.

Figure 2. Reading the message

Quantitative results

On the proficiency test, the score for all correct answers was 20 for vocabulary and 20 for grammar. According to the results shown in Table 2, the students’ English proficiency was not high (average=4.03 & 6.83 for vocabulary and grammar, respectively). In pre- and post-testing by the t-test, a significant improvement appeared in both vocabulary and grammar ($t = -6.73$, df = 34, $P = 0.000$; $t = -2.71$, df = 34, $P = 0.010$). The results confirmed that these students had developed their English from the task, and made significant improvements. Some people may argue that this improvement was temporary because the test was held immediately after the activity; however, this study showed that the proposed design could enable students to learn.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary Pre-test</td>
<td>4.03</td>
<td>2.89</td>
</tr>
<tr>
<td>Post-test</td>
<td>7.80</td>
<td>4.28</td>
</tr>
<tr>
<td>Grammar Pre-test</td>
<td>6.83</td>
<td>4.00</td>
</tr>
<tr>
<td>Post-test</td>
<td>8.34</td>
<td>4.47</td>
</tr>
</tbody>
</table>

The following table summarizes students’ responses to the questions about their attitudes, focusing on six aspects: enjoyment, interest, usefulness, effort, appropriateness of learning content, and intention to learn, which are listed as items 1 to 6, in Table 3. The responses ranged from their general attitude towards English learning, their impressions of using mobile devices to learn English and their perceptions of learning English through the tasks in this study. These results are compared in Table 3.

<table>
<thead>
<tr>
<th>Question</th>
<th>General impression</th>
<th>Impression of using mobile device</th>
<th>Task in the study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
</tr>
<tr>
<td>1 I enjoy learning English.</td>
<td>3.06</td>
<td>1.06</td>
<td>2.83</td>
</tr>
<tr>
<td>2 I am interested in learning English.</td>
<td>3.11</td>
<td>0.87</td>
<td>2.94</td>
</tr>
<tr>
<td>3 Learning English is useful to me.</td>
<td>3.97</td>
<td>0.86</td>
<td>3.23</td>
</tr>
<tr>
<td>4 I make lots of effort to learn English.</td>
<td>2.69</td>
<td>1.02</td>
<td>2.77</td>
</tr>
<tr>
<td>5 The learning content of English is appropriate to my level.</td>
<td>3.51</td>
<td>0.89</td>
<td>3.09</td>
</tr>
<tr>
<td>6 I want to learn English.</td>
<td>3.14</td>
<td>0.97</td>
<td>2.91</td>
</tr>
</tbody>
</table>
The collected data were analyzed with SPSS, and yielded a Cronbach’s Alpha coefficient of 0.854, satisfying the requirement of survey reliability. The paired sample test for significance between students’ attitudes toward English learning in general and toward the English learning task in this study are shown in items 1, 2, 4, and 6; reaching significance at \( p = 0.019, 0.000, 0.024, \) and 0.001. That is, items of enjoyment, interest, effort, and willingness to learn English improved significantly. In addition, the paired sample test between attitudes toward using mobile devices to learn English and toward the English learning tasks in this study are shown in items 1, 2, 3, 5, and 6; reaching significance at \( p = 0.002, 0.000, 0.000, 0.001, \) and 0.000.

**Discussion**

The purpose of the study was to explore the potential of mobile devices in offering advanced MALL while being technically applicable and economically affordable to language teachers. In addition, three MALL design principles and a sample practice were proposed and evaluated. The evaluation results show that, after performing the learning task, learner performance on the post-test was significantly higher than the pre-test. The MALL task can be deemed to have achieved its goal because learner’s English proficiency improved at a significant level.

This study also elucidated multiple aspects of MALL in addition to the aspects of portability and connectivity. The aspect of design was further advanced to recognize the importance of social linguistics and pragmatic usage by practicing language in authentic task-related contexts. The mobile device is an aid for creating more opportunities for communicative output, in addition to facilitating face-to-face interaction. Such design not only supports learners to have both social and digital participation in a real environment, but also confirms the interaction hypothesis and the output hypothesis in task learning. The present study echoes many researchers’ call that the design should focus on pedagogy, instead of technology, and advances the design in thinking how technology can fulfill the requirement of each knowledge field (Ting, 2010). The proposed design harmonizes the way languages are practiced with what SLA research has revealed about how they are learned; that is, understanding and practice.

This study provides an alternative solution in mobile language learning. The solution goes beyond the learning content representation types addressed by Chen, Hsieh, and Kinshuk (2008) or the context-aware content in terms of screen size and network bandwidth (Huang et al., 2008). That is, the scope of curriculum selection and instructional design is not limited to digital material within the screen of a mobile device. The scope of context becomes wider (Huang et al., 2008) and includes tasks, ambient objects, resources, peers, and a mobile device, together forming a contextualized task-based MALL. Moreover, the design is an original type of context-aware in mobile learning. Current context-aware systems (Cheverst et al., 2002; El-Bishouty et al., 2007; Liu et al., 2009) employ sophisticated technology, RFID, or QR codes to technologically synthesize digital information with the artifacts around the learners to provide context-related learning content. This study, in the field of MALL, used a deliberately designed task to synthesize the digital information and learners’ surrounding artifacts.

With regard to the issue of innovation, the mobile device is a simple communicative device without pre-crafted design for language learning, in contrast to a tailored device or even a specific system for supporting a unique learning practice. This may refer to Mulholland et al. (2005) who advocated for the spotlight browsing of web resource archives. They argued that a pre-crafted web site may require an excessive amount of effort to construct and then only provide a particular perspective on the content. By contrast, a simple database-driven site provides access to the content but little or no conceptual structure to guide its exploration. This study, therefore, added that, by employing simple mobile devices, teachers should have less technical burdens and more pedagogical discretion for designing motivating interactions among students. Such findings have been verified in this study.

Finally, the use of the proposed device involves recognizing the rapid development of commercial technology. Additionally, the device is moving us into a new era of mobile computing, promising greater variety in applications, highly improved usability, and accelerated networking; for example, the Google-led Android phone and WiMax high speed wireless network (Godwin-Jones, 2008). What this trend implies for MALL is the proposed design concept that uses advanced and mobile technology that can be sophisticated in pedagogy and simple in implementation. The sophistication means that the numerous communicative functionalities of mobile devices provide rich pedagogical design opportunities for language teachers to tailor their versatile learning activities, whose implementations are technically simple for teachers.
**Limitation and further study**

This study conducted a single experiment to evaluate the proposed design, aiming at the technical applicability of the system design. As the purpose of the study was to demonstrate how teachers can employ MALL by adding elements into tasks to enhance context authenticity, the sample design for evaluation is not large. The results cannot be generalized to other populations. In addition, as the results indicate that the whole design seemed to enhance learners’ proficiency and attitudes, the variables may need to be more clearly identified by comparing with a control group for further insights. In addition to students’ responses toward the proposed learning practice, participating teachers’ responses and their acceptance of such practice design in future teaching are required to gauge affordability. Moreover, as mobile devices advance, more peripheral equipment is added to the devices, for example, digital cameras, RFID readers, and QR code capability. These technologies may offer teachers’ more opportunities to design versatile motivating task activities in technically applicable ways as pursued in this study. Further study can explore and synthesize these contemporary technologies to provide an updated design sample.

**Conclusion**

This paper integrates the mobile device in a task-based language learning program. Drawing theoretical frameworks of second language acquisition, using teaching approaches based on task-based instruction and applying mobile technologies, a contextualized MALL was designed. This study went beyond the traditional views of mobile technology from the aspects of portability and connectivity, and adopted mobile technology as a communicative infrastructure tool without much technical effort or specialty required. Moreover, the communicative aspect of the mobile device was applied to create a motivating task within authentic contexts to support and engage students in language learning and practice. Through a practical design with implementation and evaluation, evidence shows that the participants’ improved in language skills and, more importantly, improved attitudes toward language learning. Such benefits are attributed to the coherent integration of mobile technologies with the curricula, the contexts, and the pedagogies synthesized to make task-based language learning practice original and successful.

**References**


Reducing the Impact of Inappropriate Items on Reviewable Computerized Adaptive Testing

Yung-Chin Yen, Rong-Guey Ho, Wen-Wei Liao and Li-Ju Chen

Graduate Institute of Information and Computer Education, National Taiwan Normal University, 162, He-ping East Road, Section 1, Taipei 106, Taiwan // Tel: +886 2 77343921 // Fax: +886 2 23512772 // scorpio@ice.ntnu.edu.tw // hrg@ntnu.edu.tw // ljchen@ice.ntnu.edu.tw // abard@ice.ntnu.edu.tw

(Submitted September 20, 2010; Revised May 10, 2011; Accepted August 3, 2011)

ABSTRACT

In a test, the testing score would be closer to examinee’s actual ability when careless mistakes were corrected. In CAT, however, changing the answer of one item in CAT might cause the following items no longer appropriate for estimating the examinee’s ability. These inappropriate items in a reviewable CAT might in turn introduce bias in ability estimation and decrease precision. An early proposed solution to this problem was rearrangement procedure. The purpose of this study was to implement the 4PL IRT model to reduce the estimation bias introduced by inappropriate items in reviewable CAT. The results of this study indicated that the 4PL IRT model could significantly lower the estimation bias for reviewable CAT, and, while it incorporates with rearrangement procedure, provide more accurate ability estimation. Also, the efficiency of reviewable CAT was promoted by introducing both 4PL IRT model and rearrangement procedure.

Keywords

Item response theory (IRT), Computerized adaptive testing (CAT), Reviewable CAT, Four-parameter logistical (4PL) IRT model, Rearrangement procedure

Introduction

When developing a computerized adaptive testing (CAT), an important controversy was whether examinees should be allowed to review and modify previous responses (Bowles & Pommerich, 2001; Wise & Kingsbury, 2000). In traditional paper-and-pencil (P&P) tests, most examinees took review for granted. They have been trained using the remaining time to review items and catch careless errors since they entered elementary school. In most CATs, however, opportunities for item review and answer change were far less common (Mills & Stocking, 1995; Papanastasiou & Reckase, 2007; Stone & Lunz, 1994; Vicino & Moreno, 1997; Vispoel, 1998; Wise, 1996). Since the testing score would be closer to examinee’s actual ability when careless mistakes were corrected, the prohibition of reviewing items in CAT might lead to underestimating examinees’ ability (Lunz, Bergstrom, & Wright, 1992; Vispoel, Hendrickson, & Bleiler, 2000; Waddell & Blankenship, 1994). A reviewable CAT was a CAT in which examinees were allowed to review and change the answers of previous items.

The underlying hypothesis of reviewable CAT was that after rereading or rethinking an item, the examinees might correct the careless mistake they made. This hypothesis afterwards led to the fact that even high-ability students might on occasion miss items that they should have answered correctly. However, changing the answer of one item in CAT might cause the following items no longer appropriate for estimating the examinee’s ability. These inappropriate items in a reviewable CAT might in turn introduce bias in ability estimation and decrease precision. Early proposed solutions to this problem included limiting review and rearrangement procedure (Chen, 2009; Papanastasiou, 2002).

The same situation was also seen in traditional CAT. In virtue of the underlying characteristics of the traditional item response theory (IRT) model, if a high-ability examinee misses early items carelessly in a test, the following items would be too easy to estimate his/her true ability appropriately. To cope with the underestimation problem, Barton and Lord (1981) proposed the four-parameter logistical (4PL) model allowing a high-ability student to miss an easy item without having his/her ability drastically lowered. In contrast to other well-known IRT models, however, little attention has been given to the 4PL model. The purpose of this study was to implement the 4PL IRT model to reduce the estimation bias introduced by inappropriate items in reviewable CAT. It was also hypothesized that the 4PL model would perform better in alleviating the inappropriate item problems of reviewable CAT than the rearrangement procedure, or at least, improve the performance of rearrangement procedure by incorporating with it.
4PL IRT model

According to the number of parameter describing the item, IRT model could be generally classified into three widely used categories: one-parameter logistic (1PL) model, two-parameter logistic (2PL) model, and 3PL model. In 1PL model, the probability that an examinee with ability $\theta$ could answer an item with difficulty $b$ correctly could be mathematically expressed as

$$P_{1PL}(\theta) = \frac{1}{1 + \exp[-D(\theta - b)]},$$  \hspace{2cm} (1)

The mathematical form of the 2PL model could be written as

$$P_{2PL}(\theta) = \frac{1}{1 + \exp[-Da(\theta - b)]},$$  \hspace{2cm} (2)

while new parameter $a$ was called the discrimination parameter which allowed an item to discriminate differently among the examinees (Harvey & Hammer, 1999). In both 1PL and 2PL models, the probability of passing ranges from 0 to 1 as $\theta$ goes from $-\infty$ to $\infty$. On a multiple-choice test, however, the probability of choosing the correct answer did not approach 0 even for low-ability students. Birnbaum (1968) introduced a lower asymptote to handle the situation in which examinees either guessed totally randomly or answered on the basis of their knowledge. The resulting 3PL model was

$$P_{3PL}(\theta) = c + (1 - c)P_{2PL}(\theta),$$  \hspace{2cm} (3)

where the lower asymptote $c$ represented the probability that an extremely low ability examinee would get the item correct.

In 1PL and 2PL models, the probability that a low-ability student would answer a hard item correctly should approach zero while a high-ability student should answer an easy item with probability approaching one. It was conceivable, however, this assumption might not always hold, since an examinee who knew nothing still had a chance to choose the correct answer in a multiple-choice test. Moreover, the probability might be higher for an examinee who possessed partial knowledge (Bar-Hillel, Budescu, & Attali, 2005; Burton, 2002; Gardner-Medwin & Gahan, 2003; Yen, et al., 2010). On the other hand, high-ability students who are anxious, distracted by poor testing conditions, unfamiliar with computers, careless, or misread the question, might on occasion miss items that they otherwise should have answered correctly (Hockemeyer, 2002; Rulison & Loken, 2009).

Barton and Lord (1981) introduced an upper-asymptote parameter, expressed by the Greek letter delta ($\delta$), into the 3PL model:

$$P_{4PL}(\theta) = c + (\delta - c)P_{2PL}(\theta)$$  \hspace{2cm} (4)

While $P_{2PL}(\theta)$ ranges from zero to one, $P_{4PL}(\theta)$ ranges from the lower asymptote, $c$, to the upper asymptote parameter, $\delta$, for item-specific “carelessness”. Figure 1 illustrates a typical ICC for the 4PL IRT model with $b=0$, $a=1$, $c = 0.2$ and $\delta = 0.9$.

![Figure 1. A typical ICC for the 4PL model with $b=0$, $a=1$, $c = 0.2$, and $\delta = 0.9$](image)

To evaluate whether changing the upper asymptote improved scoring on standardized tests, Barton and Lord compared the 3PL model and 4PL model under two upper-asymptote values $\delta = .99$ and $\delta = .98$ by re-estimating test
scores of four data sets: Scholastic Aptitude Test (SAT) Verbal, SAT Math, GRE Verbal, and Advanced Placement (AP) Calculus AB. The results indicated that the changes in ability estimation were too small to be of practical significance (Barton & Lord, 1981). However, it should be emphasized that this study was carried out based on the fixed response data from administered tests in which all examinees received predetermined items from the entire ability range. Hence, the next item was not dynamically selected from item bank according to examinees’ accumulated information.

To reevaluate the effect of the upper asymptote on ability estimation in a dynamically CAT environment, Rulison and Loken (2009) conducted two CAT simulation experiments to compare 3PL model with 4PL model in regard to estimation precision for high-ability students with a poor start. In the simulation an examinee (with true $\theta = 2$) missed the first two items, as Figure 2a shows, it was obvious that the initial drop was followed by a very slow ascent in $\hat{\theta}$ under 3PL model.

![Figure 2. Ability estimation in a poor-start 3PL (a) and 4PL (b) model CAT (Rulison & Loken, 2009)](image)

In contrast to Figure 2a, Figure 2b shows the trace plot for a high-ability student ($\theta=2$) who missed the first two items under 4PL model. The errors caused an initial drop almost identical to the drop in the Figure 2a, but $\hat{\theta}$ ascended faster because the upper asymptote of 0.98 discounted the early mistakes. According to Rulison and Loken’s study, using 4PL model ($\delta = 0.98$) could lower estimation bias for high-ability students with a poor start. In other words, 4PL IRT model proposed examinees an opportunity to recover from unreasonable responses in CAT.

**Reviewable CAT**

The term “item review” in testing contexts referred to administrative rules that allowed examinees to change their responses to previously answered items (Vispoel, Rocklin, Wang, & Bleiler, 1999). The opportunity to review items and change answers was usually important to examinees for a variety of reasons (Gershon & Bergstrom, 1995; Harvil & Davis III, 1997; Heidenberg & Layne, 2000; McMorris & Weideman, 1986; Mills & Stocking, 1995; Shatz & Best, 1987):

- Examinees might remember some facts that he/she has not previously recalled, and recognize that they have answered an item incorrectly later in the test.
- Examinees might rethink and conceptualize a better answer for a previous item.
- Review might also be the result of an item occurring at one point in a test cuing the correct answer to a previous item.
- Finally, examinees might want to use the remaining time either to reread the items to avoid careless error, or just to guess another answer to previous questions.

These reasons for choosing to review items could be divided into two major categories: legitimate and illegitimate ones (Wise, 1996). Legitimate reasons were those in which examinees change incorrect to correct answers due to knowledge possessed before the test. This was considered good practice since the final score would reflect the
examinee’s ability more accurately. In turn, the validity of the test increased. Illegitimate reasons for those changing answers included the cases in which examinees corrected an incorrect response due to test-wisdom. In this case, the validity of the test decreased (Papanastasiou, 2005). This was the main reason why most CATs did not allow examinees to review items.

Due to the complexity and difficulty of implementation, reviewable CAT was in fact quite rare in practice (Parshall, Kalhn, & Davey, 2002). However, Vispoel et al. (2000) and Papanastasiou (2002) still proposed two algorithms of a reviewable CAT procedure, respectively. These solutions were described in the following two sections.

**Limiting answer review and change procedure**

Vispoel et al. (2000) proposed the limiting answer review and change procedure that allowed reviewing and changing items within successive m-item blocks. Compared with the traditional CAT, the test items were grouped into n blocks. Figure 3 shows the flow chart of limiting answer review procedure of CAT. In this procedure, examinees were only allowed to review and change answers within the recent block. If an examinee was answering the items in block \( j \), he/she was not allowed to review the items in the previous blocks.

![Figure 3. The procedure of limiting reviewable CAT](image)

To verify the performance of the limiting review procedure, an experiment randomly assigning subjects to four CAT conditions was conducted: no review, review of items in 4 blocks of 10 items each, review of items in 8 blocks of 5 items each, or review of all items at the end (Vispoel et al., 2000). In all review conditions, examinees were satisfied with their review options, even when the block size was small. Though the average ability of all review conditions was higher than that of the no review condition, ability estimation did not seem to be related to the various review conditions. As the block size increased, examinees tended to spend more time on the test while there was only 6% difference in testing time between the no review and review in 5-item blocks conditions. The result suggested that block review might function just as well as full review while we reduce item selection problem associated with review.

There were two advantages for using the limiting answer review and change procedure. Firstly, the problem of Wainer strategy could be overcome. That is, the examinees' cheating strategy would not have much effect when they were allowed to review items in a block only (Stocking, 1997; Vispoel et al., 2000). Secondly, there were no significant difference in the accuracy of ability estimation between the limiting review and the no review procedure (Vispoel et al., 2000). This suggested that an examinee could still gain accurate ability estimation by using a reviewable CAT. However, the item administration sequence might become unreasonable after they change answers in this procedure. For example, examinees might correctly change the answers for more difficult items, but such
change would lead to unreasonable item-administration sequence in which some easier items followed a correctly changed, difficult item. These unreasonable response patterns might lead to serious ability estimation bias and decrease testing precision. The rearrangement procedure was proposed to cope with this problem by rearranging the response patterns and re-estimating examinees’ ability after answers were changed.

Rearrangement procedure

Papanastasiou (2002) proposed the rearrangement procedure that rearranged and skipped certain items and could better estimate the examinees’ abilities. For example, the rearrangement procedure allowed examinees to change up to five of their answers after they finished 30 items in the allotted testing time. After examinees revised their answers, the examinees’ final scores were calculated.

Compared with the traditional CAT, three types of answer changing (Change I→I, Change I→C, and Change C→I) caused the rearrangement procedure in ability estimation. Change I→I involved changing answers from incorrect to incorrect and it would make no difference in ability estimation between the traditional CAT and the rearrangement procedure. The second type (Change I→C) involved changing answers from incorrect to correct and it would result in item skipping in the rearrangement procedure. For instance, if Change I→C was made to item \( i \), the ability estimation \( \hat{\theta}_i \) would be changed to \( \hat{\theta}'_i \). This change in ability estimation would probably make the item \( i+1 \) not the most informative item for the ability \( \hat{\theta}'_i \) since it would be easier and was originally targeted at ability \( \hat{\theta}_i \) rather than \( \hat{\theta}'_i \). To solve this problem, the ability estimation would skip to item \( X \) (e.g., item \( i+k \), with \( 1<k<4 \)) that was answered incorrectly if it was more difficult. It was hypothesized that the new item \( X \) would be more similar to the item that would have been administered after item \( i \). The third type (Change C→I) was to change answers from correct to incorrect and it would also result in item skipping in the rearrangement procedure. For example, if Change C→I was made to item \( j \), the item \( j+1 \) would be ignored in the ability estimation procedure since it would be targeted at a higher ability level than \( \hat{\theta}_j \). Therefore, the ability estimation would skip to item \( Y \) (e.g., item \( j+k \), where \( 1<k<4 \)) if that was the first easier and correctly answered item after item \( j \). So it was hypothesized that the ability estimation
would be more accurate if items \( j+1 \) through \( j+k-1 \) were ignored from the estimation procedure, and item \( j+k \) was used after the item whose answer was changed.

As Figure 4a shows, a Change \( I \rightarrow C \) (incorrect-to-correct) was made to item 2 of a reviewable CAT. As a result, item 3 became an inappropriate item after the change of this answer since it was selected from the item bank on the premise that the examinee answer item 2 was wrong. In this case, item 3 was skipped because it was targeted at a lower ability level than \( \hat{\theta}_j' \). The algorithm therefore jumped to item 4 since that was the first more difficult item answered incorrectly that came after item 3 (see Figure 4b).

The underlying hypothesis of reviewable CAT led to the fact that high-ability students might on occasion miss items that they should have answered correctly. However, almost all previous reviewable studies were carried out based on traditional CAT that assumed that a high-ability student should answer an easy question with probability approaching one. As descried above, the underlying hypothesis of reviewable CAT was consistent with the principle of 4PL model. Besides, the 4PL model might propose examinee an opportunity to recover from the inappropriate responses which introduced by reviewing and changing answer in reviewable CAT. In the present study, therefore, the effect of 4PL model and the rearrangement review solution on reducing estimation bias was investigated.

**Methodology**

In this study, a simulated experiment was conducted to evaluate the effect of implementing the 4PL model on reducing the impact of inappropriate items on reviewable CAT. The participant demography, item bank characteristics, and procedure are described as follows.

**Participants**

A group of 13,000 examinees were simulated for this study (1,000 examinees for thirteen equally spaced \( \theta \) levels). The \( \theta \)-level groupings ranged from -3.0 to 3.0 at equally spaced intervals of 0.5.

**Item bank**

The simulated item bank with 250 items was generated according to the specifications proposed by Papanastasiou and Reckase (2007). To determine the item pool characteristic for their simulation study, Papanastasiou and Reckase reviewed psychometric literature to obtain information on the distributions of item parameters of real item pools. Table 1 describes the targeted distributional characteristics of the item pool created for this simulation. As for the upper asymptote, a test-wide \( \delta \) with value of 1 for 3PL-based CAT and 0.98 for 4PL-based CAT was determined, respectively. By designing item bank of 3PL- and 4PL-based CAT in this way, the only difference between these two models would be the value of upper asymptote, and the results would be correspondent with the intention of this study.

**Table 1.** Distributional Characteristics of the Item Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type of distribution</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>Uniform</td>
<td>0.00</td>
<td>2.00</td>
<td>-3.50</td>
<td>3.50</td>
</tr>
<tr>
<td>a</td>
<td>Log normal</td>
<td>1.10</td>
<td>0.25</td>
<td>0.45</td>
<td>2.30</td>
</tr>
<tr>
<td>c</td>
<td>Uniform</td>
<td>0.17</td>
<td>0.10</td>
<td>0.00</td>
<td>0.35</td>
</tr>
</tbody>
</table>

**Simulation procedure**

Four versions of 30-item, fixed-length CAT were developed for this simulation experiment (as Figure 5 shows). All these four CATs were reviewable CAT with 6 blocks of 5 items, and the reviewing and changing answer were allowed only within block. The first two CATs (R3CAT and R4CAT) were conducted to compare the performance
of 3PL- and 4PL-based CAT on precision and efficiency. The third one, RR3CAT, was a reviewable CAT with rearrangement procedure while the RR4CAT was a 4PL-based reviewable CAT implementing rearrangement procedure. These two CATs were administered to investigate whether the precision and efficiency of reviewable CAT with rearrangement procedure would be improved by implementing the 4PL IRT model.

1. **R3CAT**: A reviewable CAT based on 3PL IRT model.
2. **R4CAT**: A reviewable CAT based on 4PL IRT model.

In all four CATs, the first test item started at an item with a middle difficulty. Bayesian expected a posteriori (EAP) estimation was used to estimate examinees abilities in all four CATs while items were selected using a simple maximum-information criterion. Based on the item parameters and simulated $\theta$s, the probability ($p$-value) of an examinee answering an item correctly according to the corresponding IRT model was calculated. This $p$-value then was compared to a randomly generated number ($x$) from a uniform distribution $U(0,1)$. If the $p$-value is larger than or equal to the uniform random number, the simulee got a correct response; otherwise, an incorrect response was obtained for the item. Referring to the study of Rulison and Loken (2009), the Fisher’s information for R3CAT and RR3CAT was given by

$$I_j(\theta) = \frac{(1.7a_j)^2(1-c_j)}{(c_j + e^{-1.7a_j(\theta-b_j)})(1+e^{-1.7a_j(\theta-b_j)})^2}$$  \hspace{1cm} (5)$$

and Fisher’s information function for the R4CAT and RR4CAT was given by

$$I_j(\theta) = \frac{(1.7a_j)^2(\delta-c_j)^2}{(c_j + e^{1.7a_j(\theta-b_j)})(1-c_j + (1-\delta)e^{1.7a_j(\theta-b_j)})(1+e^{-1.7a_j(\theta-b_j)})^2}$$  \hspace{1cm} (6)$$

where

$I_j(\theta)$ was the item information at $\theta = \hat{\theta}$ for item $j$,

$a_j$ was the item discrimination parameter for item $j$,

$b_j$ was the difficulty parameter for item $j$,

$c_j$ was the lower asymptote parameter for item $j$, and

$\delta_j$ was the upper asymptote parameter for item $j$.

As five items within each block were answered, the reviewing and changing answer procedure took place. One of three situations might occur in each rearrangement procedure: 1. If the simulee had a 0.80 or higher probability of answering these items correctly ($p$-value $\geq 0.80$), those answers would be changed from incorrect to correct (Change $I \rightarrow C$); 2. If the simulee only had a 0.33 or lower probability of answering correctly ($p$-value $\leq 0.33$), the answer would be changed from correct to incorrect (Change $C \rightarrow I$) during the revising and changing procedure; 3. For those simulees whose probability to correctly answer an item was from 0.47 to 0.53 ($0.47 \leq p$-value $\leq 0.53$), the probability of answering item correctly (Change $I \rightarrow C$) would be 0.72 and the probability of answering item incorrectly (Change $I \rightarrow I$) would be 0.28. Figure 5 shows the flowchart of four reviewable CATs.
C→I or Change I→I) would be 0.28. These probabilities described above were determined according to Papanastasiou’s (2005) study.

**Results**

By comparing the estimation bias (Equation 7), median absolute deviation (MAD) (Equation 8), and standard error of estimation (SE) of R3CAT, R4CAT, RR3CAT, and RR4CAT, the precisions of 3PL and 4PL models on reviewable CAT were investigated. On the other hand, by comparing the number of items needed to reach certain SE levels, the efficiency of the four CATs each was evaluated.

\[
\text{Bias}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta_i) \quad (7)
\]

\[
\text{MAD}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} |\hat{\theta}_i - \theta_i| \quad (8)
\]

Table 2 shows that the descriptive statistics of 13,000 simulees’ three change types; the conditions of two regular reviewable CATs (R3CAT and R4CAT) were close to the results of meta-analysis by Waddell and Blankenship (1994). One interesting finding revealed that the number of Change I→C (change from incorrect to correct) in reviewable CAT with rearrangement procedure (RR3CAT and RR4CAT) was smaller than that in traditional reviewable CAT (R3CAT and R4CAT). In a rearrangement procedure, if Change I→C was made to item \(i\), the following easier items might be skipped since the ability re-estimation probably make these items (item \(i\) to item \(i+k\), with \(1<k<4\)) not appropriate for the newly estimated ability. These skipped items, however, might be the potential targets for Change I→C. This explained the lower rate of Change I→C in RR3CAT and RR4CAT. Though those inappropriate harder items might also be skipped in Change C→I, the higher rate of Change I→C indicated that more inappropriate easier items were skipped than inappropriate harder items in rearrangement procedure. This also explained the higher rate of Change C→I in RR3CAT and RR4CAT compared with that in R3CAT and R4CAT.

<table>
<thead>
<tr>
<th>Change type</th>
<th>CAT type</th>
<th>R3CAT (%)</th>
<th>R4CAT (%)</th>
<th>RR3CAT (%)</th>
<th>RR4CAT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change I→I</td>
<td></td>
<td>21.59</td>
<td>24.75</td>
<td>20.85</td>
<td>22.47</td>
</tr>
<tr>
<td>Change I→C</td>
<td></td>
<td>67.55</td>
<td>64.17</td>
<td>52.38</td>
<td>49.12</td>
</tr>
<tr>
<td>Change C→I</td>
<td></td>
<td>10.87</td>
<td>11.08</td>
<td>26.76</td>
<td>28.41</td>
</tr>
</tbody>
</table>

To exactly investigate the effect of 4PL model and rearrangement procedure on reviewable CAT, it was essential to distinguish reviewed items and changed answers. The reviewing behaviors included Change I→I, Change I→C, and Change C→I. The changed answers meant those changes of responses that would cause the re-estimation of ability (i.e., Change I→C + Change C→I). Table 3 presents the percent of reviewed items, changed answers, and examinees who changed at least one item during the CAT administration. Since our main concern in this experiment was for those examinees who changed at least one item in the simulation, simulees who never changed answer during the simulation were excluded from the following evaluation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>CAT</th>
<th>R3CAT (%)</th>
<th>R4CAT (%)</th>
<th>RR3CAT (%)</th>
<th>RR4CAT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewed items</td>
<td></td>
<td>14.03</td>
<td>13.26</td>
<td>12.36</td>
<td>12.52</td>
</tr>
<tr>
<td>Changed answers</td>
<td></td>
<td>11.00</td>
<td>9.98</td>
<td>9.79</td>
<td>9.71</td>
</tr>
<tr>
<td>Simulees changed at least one item</td>
<td></td>
<td>96.38</td>
<td>95.02</td>
<td>97.55</td>
<td>97.35</td>
</tr>
</tbody>
</table>

Table 3. Percent of changing behavior (Number of Responses=3,900,000)
The precision of four solutions for reviewable CAT

To investigate the performance of 4PL-based IRT model and rearrangement procedure in regard to diminishing estimation bias caused by inappropriate items in reviewable CAT, the estimation bias of four CATs were computed. Figure 6 indicates that the ability underestimation of 3PL-based CAT (R3CAT and RR3CAT) was alleviated in 4PL-based CAT. The bias for the 4PL-based CATs tended to be smaller than that of 3PL-based CAT, except for the top and bottom ability levels.

Figure 6. Estimation bias of four CATs

To further evaluate the degree of estimation bias of the four CATs, the estimation MADs of these four CATs were calculated and the ANOVA of MAD was conducted. The value of MAD indicated the distance between the examinees’ estimated ability and true ability. The lower the value of MAD was, the more precise the ability estimation was. As Figure 7 shows, the estimation MADs of R4CAT were lower than those of R3CAT across all ability levels. The performance of RR4CAT was also better than that of R3CAT except for some ability levels ($\theta=-3.0, -0.5, \text{and } 3.0$). On the other hand, the performance of reviewable CAT with rearrangement procedure (RR3CAT) was similar to that of R3CAT. The lower MAD indicated that the 4PL-based IRT model was a better solution to reviewable CAT.

Figure 7. Estimation MAD of four CATs

Since the purpose of this study was to evaluate the performance of 4PL-based IRT and rearrangement procedure by comparing the four CATs, 390 simulees (30 simulees in each ability level) who changed at least one item in each CAT were randomly selected and the repeated measures ANOVA of MAD was conducted. This sampling method was applied to the following repeated measures ANOVA in this experiment. The result of the Mauchly’s Test of sphericity indicated that there was no significant difference among the four sets of MAD. A summary of the results of the repeated measures ANOVA is shown in Table 4. As Table 4 shows, though the precision of reviewable was
improved by rearrangement procedure (RR3CAT and RR4CAT), the difference among R3CAT, RR3CAT, and RR4CAT was not significant. According to the results of repeated measures ANOVA, the precision improvement by incorporating 4PL IRT model with rearrangement procedure (RR4CAT) was not statistically significant. Compared with these three CATs, R4CAT significantly improved the precision of ability estimation for reviewable CAT.

Table 4. Repeated Measures ANOVA of MAD (n=390)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Pairwise comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT type</td>
<td>1.09</td>
<td>3</td>
<td>0.36</td>
<td>2.64*</td>
<td>R4CAT &lt; RR4CAT, RR3CAT, R3CAT</td>
</tr>
<tr>
<td>Error</td>
<td>160.47</td>
<td>1167</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.

To further investigate the difference of estimation precision between these four CATs, the estimation SE of these four CATs was evaluated. Figure 8 indicates that SEs of 4PL-based CAT (R4CAT and RR4CAT) tended to be smaller than that of 3PL-based CAT (R3CAT and RR3CAT) across most ability levels, except for the very low ability level (θ=−3.0). The precision improvement introduced by rearrangement procedure was also obvious since the SE of RR3CAT was smaller than that of R3CAT and the SE of RR4CAT was smaller than that of R4CAT across all ability levels. However, according to the results of repeated measures ANOVA, the precision improvement introduced by 4PL IRT model was more significant.

Table 5. Repeated Measures ANOVA of SE (n=390)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Pairwise comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT type</td>
<td>0.03</td>
<td>3</td>
<td>0.01</td>
<td>2.64*</td>
<td>RR4CAT&lt; R4CAT&lt; RR3CAT&lt; R3CAT</td>
</tr>
<tr>
<td>Error</td>
<td>0.11</td>
<td>1167</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05.

The efficiency

The efficiency of CAT was evaluated by comparing all four versions of CAT within the average number of items required for each examinee’s ability estimation to reach or surpass a fixed level of precision. Based on a set of stopping criteria including SE reaching to 0.45, 0.40, 0.35, and 0.30, the required numbers of items for R3CAT, R4CAT, RR3CAT, and RR4CAT were compared. Figure 9a depicts the required number in each ability levels for four CATs while SE reaches to 0.45. According to the test length plots, the efficiency of reviewable CAT was improved by introducing the 4PL-based IRT model and rearrangement procedure since the required items of R4CAT,
RR3CAT, and RR4CAT were all smaller than that of R3CAT. The efficiency of R4CAT was superior to RR3CAT at the middle and higher ability levels ($\theta \geq 0$) while the RR3CAT performed better at the rest of ability levels. The efficiency of RR4CAT was superior to that of other three CATs across all ability levels.

Figure 9. Mean of test length for four CATs

Figure 9b describes the mean number of items in each ability level for four versions of CAT while SE reaching to 0.40. Like the situation in the previous criterion, the efficiency of reviewable CAT was improved by introducing the 4PL-based IRT model and rearrangement procedure. Compared to RR3CAT, the required number of items for R4CAT was smaller except for the ability levels ranging from -2.0 to 0.5. RR4CAT maintained the best performance of efficiency across all ability levels. As Figure 9c shows, the R4CAT was more efficient than RR3CAT at the higher and lower ability levels. This result was similar to that of previous stopping criterion. While the stopping criterion was set to SE less than or equal to 0.30 (see Figure 9d), four CATs’ performance of efficiency was consistent with those of previous two stopping criteria. The efficiency of RR4CAT was identically superior to other three CATs across all ability levels.

To investigate the differences in efficiency among these four CATs, repeated measures ANOVA of test length based on four preset stopping criteria was conducted. Though the variances in the differences between the four sets of test length did not meet the assumption of sphericity according to the result of Mauchly’s test of sphericity, the adjusted $F$ was the same after the Greenhouse-Geisser adjustment was consulted. Results of repeated measures ANOVA is provided in Table 6.

In summary, there was a significant difference in efficiency among these four CATs. Both 4PL-based IRT model and rearrangement procedure improved the efficiency of reviewable CAT significantly. Though the average required number of item for R4CAT was smaller than that for RR3CAT across all stopping criteria, the differences were not statistically different. The efficiency of reviewable CAT was significantly improved while the 4PL IRT model incorporated with rearrangement procedure (RR4CAT).
Table 6. Repeated Measures ANOVA of Test Length (n=390 in each CAT type)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Pairwise Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE ≤ 0.45</td>
<td>CAT type</td>
<td>98.13</td>
<td>3</td>
<td>32.71</td>
<td>40.37***</td>
<td>RR4CAT&lt;R4CAT, RR3CAT&lt;R3CAT</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>945.63</td>
<td>1167</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE ≤ 0.40</td>
<td>CAT type</td>
<td>104.86</td>
<td>3</td>
<td>34.95</td>
<td>31.93***</td>
<td>RR4CAT&lt;R4CAT, RR3CAT&lt;R3CAT</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>1277.39</td>
<td>1167</td>
<td>1.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE ≤ 0.35</td>
<td>CAT type</td>
<td>236.18</td>
<td>3</td>
<td>78.73</td>
<td>33.68***</td>
<td>RR4CAT&lt;R4CAT, RR3CAT&lt;R3CAT</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>2727.82</td>
<td>1167</td>
<td>2.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE ≤ 0.30</td>
<td>CAT type</td>
<td>694.11</td>
<td>3</td>
<td>231.70</td>
<td>53.79***</td>
<td>RR4CAT&lt;R4CAT, RR3CAT&lt;R3CAT</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>5026.64</td>
<td>1167</td>
<td>4.31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***p<.001.

Conclusion

In a test, the testing score can be closer to examinee’s actual ability when careless mistakes are corrected. The rearrangement procedure proposed by Papanastasiou (2002) offers examinees opportunity to review and change administered items in a CAT. By limiting the number of reviewing items, the rearrangement can not only avoid cheating strategy caused by reviewing, but also require no extra testing time. Moreover, the rereading and rethink process is important since taking a test is not just a passive mechanism for assessing student. Taking a test actually helps student learn, and it may work better than a number of other techniques.

In CAT, however, changing the answer of one item in CAT might cause the following items no longer appropriate for estimating the examinee’s ability. These inappropriate items in a reviewable CAT might in turn introduce bias in ability estimation and decrease precision. This study implemented the 4PL IRT model as a solution to the problem of estimation bias introduced by inappropriate items in reviewable CAT. The simulation result indicated that the 4PL IRT model could significantly lower the estimation bias for reviewable CAT, and provide more accurate ability estimation while incorporating with rearrangement procedure. Also, the efficiency of reviewable CAT was promoted by introducing both 4PL IRT model and rearrangement procedure.

Acknowledgements

This work was supported in part by the Science Education Division of Taiwan National Science Council under the Grant No. NSC 99-2511-S-003-010-MY2. The project principal investigator, Rong-Guey Ho, is also the corresponding author of this article. The authors would like to thank the NSC of Taiwan for financial support.

References


Modelling Digital Natives’ International Collaboration: Finnish-Korean Experiences of Environmental Education

Irja Leppisaari¹ and Okhwa Lee²*

¹AVERKO, Central Ostrobothnia University of Applied Sciences, Talonpojankatu 2, 67100, Kokkola, Finland // ²Dept. of Computer Education, Chungbuk National University, 410 Sungbong-ro, Heungdok-ku, Cheongju city, Chungbuk, Korea 361-763 // irja.leppisaari@cou.fi // ohlee@cbnu.ac.k

*Corresponding author

(Submitted October 4, 2010; Revised February 10, 2011; Accepted April 10, 2011)

ABSTRACT

A new generation of young learners often described as digital native school children are attitudinally and technically equipped to employ social media as a social process in learning. However, few international virtual learning projects have been implemented and researched. This article examines a trial which aimed to combine viable technology with future pedagogic solutions for primary students from Korea and Finland and create an international collaboration model in virtual learning for environmental education. The results show various challenges of the operational model and suggest effective implementation strategies. The challenges were organisational, language, technical and collaboration barriers. The operational model illustrates possibilities of implementing cyber space pedagogy, visualization of knowledge using technology, cyber spaces for collaboration, and the motivational impetus provided by the model. This pilot study demonstrates the need to increase greater interactivity between teachers from the partner countries during the planning phase and provide more authentic interaction for inter-learner dialogue.

Keywords

International collaboration model, Environmental education, ICT in education, Finland, Korea

Introduction

As the world becomes more inter-connected and globalized, people are asked to be aware of multi-cultural differences, and global education is particularly needed for 21st century students (Cisco, 2010). Schools are often asked to provide education for global citizenship (Kaijola & Melén-Paaso, 2007; O’Neill, 2006). Information communication technology enables communities to move beyond geographical boundaries and provides a vehicle for people around the world to interact and learn together seamlessly. Concurrently learners in international collaboration on the web are required to face and overcome numerous barriers, e.g. linguistic, temporal and cultural boundaries (Walker & Creanor, 2011).

Direct interaction and communication with students from other cultures can be one of the most effective ways to understand and learn about cultural differences. Global environmental issues can be an appropriate means of introducing global citizenship to students. Social software, like Web 2.0 enables students to collaborate through computer-mediated communication and to form learning communities in which they construct and share knowledge. Learning communities emerge when students share common interests (Jonassen, Howland, Marra & Chrismond, 2008; Hakkarainen, Paavola & Paavola, 2004). Since ICT is already developed enough to easily provide students with international interaction opportunities in teaching and learning, schools should not hesitate to offer international collaboration using ICT, which seems to be a promising means of achieving this goal (O’Neill, 2006).

ICT has permeated daily life extensively, and Finland and Korea are known as highly wired countries. Not only are the societies technically equipped, but schools are also. Although great differences continue to exist in resources between schools, extensive national training and school facilities have ensured adequate proficiency of teachers’ technical skills and ability to employ ICT in education. However, schools still face challenges – the use of technology as a learning tool has not increased (Niemi & Kumpulainen, 2008; Kankaanranta & Puhakka, 2008). Developing school practices to reflect contemporary learning concepts powerfully impacts technologic needs in teaching (Niemi & Kumpulainen, 2008). Infrastructure of ICT in Korean and Finnish schools has been well established but pedagogic implementation of ICT in education has not been adopted well.
Digital native school children challenge pedagogy

Digital natives and net generation are two common phrases used to describe the generation of people born between the early 1990s and early 2010s. They have grown up with digital technology and are the first generation to be bathed in bits (Tapscott, 2009), creating the assumption that this generation has a natural aptitude and high skill levels for using new technology (Jones, Ramanau, Cross & Healing, 2010). Yet Jones and Healing (2010) approach the ‘Net Generation’ or ‘Digital Natives’ assumption at macro and micro levels with reservation. They claim that the net generation trend should be treated with meso level like class or program activities, as students in advanced industrial countries are far from homogenous in their response to new technology. The arguments raised by Jones and Healing (2010) underpin the use of the digital natives or net generation concepts in this paper.

An educational challenge for the global world is to provide learning experiences of authentic operational cultures that support interaction between individuals and groups (Vähähyyppä, 2010). Being exposed to a different culture can deliver an optimal learning experience, yet not all have access to such opportunities. Thus the experience of different cultures through cyber space becomes a feasible solution to meet this educational challenge. Technology becomes a vital environment for global citizenship education. Despite the importance of ICT for global education, virtual international projects between elementary schools have been little researched (O'Neill, 2006; Korkala, 2009).

A critical question is how viable technology can be integrated to future pedagogic solutions to create authentic learning experience. Green education, also known as environment study, is a strongly emerging theme in Korea (MEST, 2010). The United Nations has declared 2005–2015 the Decade of Education for Sustainable Development. Basic education endeavors to produce skills for environmental protection that take into consideration socio-cultural and technologic-economic objectives of sustainable development (Tapio, Kohl, Tikkannen & Salonen, 2007). Environment study is often used in global education as it requires international collaboration and an understanding of being connected globally regardless of differences in culture, race, and nationality.

The pilot examined in this article is a Korean and Finnish educator initiated environmental education project at elementary school level. The Korean and Finnish education systems are quite similar: 6-3-3 (six years of elementary school, 3 years of junior high school, 3 years of senior high school). Class size in Korea is about 35 students, while in Finland class groups during basic education comprise 20-25 learners. English lessons in both countries begin in third grade (9-10 year-old) at the latest. The project involved five classes in Finland (approximately 120 students) from the same elementary school and three classes from three different schools in Korea (approximately 100 students). Classes were selected arbitrarily in Finland, and in Korea schools with teachers belonging to the Global Education Foundation (GEF) (www.globaleducation.or.kr) were selected. The Finnish students were 10–12-year-olds (grades 4–6) and the Korean students 11–12-year-olds (grades 5 and 6). The project was conducted during the fall semester (Sept. - Dec.) in 2009. The pilot aimed to create an international cooperation model for elementary school environmental education. The project stimulus came from a Korean online education partner. The content used for the international collaboration was waste recycling and instructional materials came from the Korean research team.

Research

The Korean-Finnish environmental education project investigated the following research questions:

• How does a cyber space function as an environmental education meeting place for today’s digital native school children? What challenges and strengths emerged in implementing technology?

• What kind of operational model would be meaningful in future international virtual collaboration at elementary school level? Which cultural factors are influential and how?

The research data primarily comprises log information, records and content saved in the Edu2.0 platform: discussions, interactive processes, produced material, and emails. Additionally, the activities of one Finnish class were observed during the project period. The data collected were used to answer the first research question. Finnish teachers (5) and one principal were group interviewed at the end of the project in January 2010. Data from the four Korean teachers was collected through in-depth interviews. Qualitative data from interviews were used to develop the operational model (Cohen, Manion & Morrison, 2000). The qualitative analysis used content categorization.
Selection of the virtual tool

The tool ‘Edu2.0’ (http://www.edu20.org/) (see Figure 1) was selected after a careful analysis of various alternatives as it contained features suitable for the project objectives such as user-friendliness, learning management system, interaction, and functions of sharing knowledge. The discussion forum was the main interaction area in this project. Students could, for example, send and receive messages, interact with each other, post pictures and videos in the cyber space, and read and comment on each other’s productions.

The Edu2.0 learning platform is open software which does not require any software installations or server environments. This service, administered from San Francisco, USA, was established in 2006 and has over 150,000 users globally. Environment creation was initiated by the Korean partners, but was easily adaptable for any changes as the process advanced. Management of the Edu2.0 platform was facilitated by the Korean research team. Visual sharing of knowledge was critical in the project so that in addition to Edu2.0, digital cameras, YouTube and Photoshop software and other image editing tools were also used. Private data was only transmitted through Edu2.0, as its SSL login and other security measures were deemed adequate to protect material.

Figure 1. Edu2.0 website

Orientation and initiation

Teacher orientation programs for the entire project and its pedagogic idea were held separately in both countries. In Finland, the operational phase of the project for the 5 participating teachers began at the end of August 2009 just before school resumed. The principal lecturer of online pedagogy at a Finnish university of applied sciences and the principal of the participating elementary school conducted the orientation of the pilot’s operational concept and pedagogic operational model. Issues related to class and sub-group virtual collaboration were considered and viable solutions sought. The orientation also included training in the use of digital tools, familiarization with the joint virtual working environment, and information about the use of the instructional materials for ‘recycled materials’. A Korean IT specialist compiled an Edu2.0 manual, which was translated into Finnish.

Previous studies have shown that organization and availability of continuous technical support is a critical success factor for international school learning collaborations (O’Neill, 2006; Lee, 2009). The Finnish partner resourced 4 hours of in-class IT support each week. A course environment was created for pairs of students in Edu2.0. Small groups, whose task was to create and handle together environmental education knowledge, were created in the course discussion forum.
The Korean team also conducted an orientation program for the four Korean teachers more or less in the same way as in Finland and a team of instructional developers designed instructional materials. In Finland the principal’s role in initiating the project was significant. According to Korean practice and protocol, teachers manage projects of this scale. However, it was obvious that activities went beyond a classroom teacher’s responsibility, and help was required at school level in the form of translation support, more technical support, and field trips to waste recycling facilities. Class selection was done voluntarily through the GEF teachers’ association.

The project’s frame

The learning tasks of the project phases were: 1) Ice-breaking, 2) Collecting information about waste recycling in their environment, and 3) Presenting creative ideas for recycling waste. A Korean researcher designed the learning tasks, intended as guidelines, and teachers were given freedom to adapt, fine-tune and apply these to meet their needs. The learning tasks are introduced in more detail below:

Ice-breaking

Multimedia affordances were utilized in ice-breaking. The Internet generation is accustomed to information presented on screens and in diverse forms: images, videos, music (Vähähyppä, 2010). Students brought pictures of themselves, their family, hobbies and homeland into the closed learning environment, and discussed these themes. Favorite music was also added to the Edu2.0 platform.

Figure 2. Observations on waste recycling made by South Korean students

Figure 3. Observations on waste recycling at a Finnish school
Environmental themes emerged strongly in these introductions, especially in narratives of one’s homeland. One Korean student wrote: “Hello! my name is XX! my favorite food is kimchi.... My country Korea is not clean but we are separate the trash, wow! It is real! If you come to Seoul, you’ll feel very happy and you’ll love my city. Seoul is peaceful city, lovely city and famous city. All student and I am endeavor for earth because if the earth destruction, it will be bad to us.”

**Collecting information about waste recycling in their environment**

Themes investigated in this project phase included waste generated at home or school. Students used digital cameras to photograph their school neighborhood. The photos were utilized in gathering knowledge on waste recycling as required by the learning task especially in Korea, but also partly in Finland (see Figure 2 and Figure 3). The photos were eloquent illustrations of student observations.

**Presenting creative ideas for recycling waste**

In the final phase of the project, students used creativity to recycle waste—in fact they produced pieces of art. The aim was a reciprocal presentation and sharing of product ideas among students through the cyber space. This phase of using visuals was purposely included to minimize the language barrier and maximize the dramatic results of students’ creative application of knowledge. We expected visuals to compensate for language limitations.

---

**Figure 4.** Pictures of handicrafts students made of reused waste

Finnish students made, for example, furniture coasters for their desks from milk cartons. Finns drink a lot of milk, generating a lot of milk carton waste during school meals. In the second phase, students interviewed school catering staff about milk carton waste. They then calculated how much carton waste is produced daily. The class needed 160
milk cartons to make their coasters. Other arts and craft were made from recycled products, such as candle lanterns from glass jars, and jewellery fashioned from paper, buttons and beads. Pictures of the products (see Figure 4) were posted in the Edu2.0 platform for others to see.

Challenges in the operational model

Organization barriers

Several barriers and boundaries were identified during the project. Due to numerous organizational, scheduling, language and interaction challenges, project objectives were only partially achieved. In her research on the internationality and various operational forms of Finnish basic education, Korkala (2009) stresses the fact that participation in international projects requires time and resources. These projects should not disrupt core work and burden staff; rather they should be integrated into curriculum activity and be a continuous process tightly coordinated within everyday school work (Korkala, 2009).

This project’s commencement timetable was tight and it began rather spontaneously. The original objective was to create an operational model with a large number of participants. Timetable mismatches existed due to organizational difficulties and school vacations. For example, the number of classes participating from South Korea was less than originally planned and therefore only one Finnish class was matched with a Korean partner class. In Finland, separate resources were not allocated to the project, which in practice determined the project’s importance amongst the heavy school workload. The time assigned for the use of computers was not sufficient for participating students to write, read, discuss and respond to counterpart students’ postings. International cooperation should be scheduled before the school fiscal calendar is decided, which in Korea is in February. Thus, as this project started in the fall semester, some schools that planned to participate were unable to fit the project into the school curriculum smoothly. Only one out of three schools was able to participate from Korea. This was also due to lack of school leadership where principals of the three schools showed little enthusiasm at the beginning of the project. School leadership is essential for the success of these kinds of international projects.

Difficulties in the roles, responsibilities and flow of information emerged due to a lack of direct communication between teachers. Teachers from Finland and Korea lacked opportunities to build trust between each other before the project started. But trust is a key factor for successful virtual learning community activity (Lewis & Allan, 2005). Communication primarily occurred through “intermediaries,” that is, between project initiators. The operational model’s core difficulty in this experiment was the creation of a concrete collaboration model among classes and groups. This kind of international collaboration requires well defined support activities, which include among other aspects, recognition, priority of incentives and cultural acceptance.

Language barriers

One significant boundary was language. English was to be the primary communication medium as one project aim was to provide authentic opportunities for students to use English. However, it was quickly apparent that using English to complete class tasks was too challenging for ten-year-olds, even with support from the English teacher. Classes were dissimilar and differences of language proficiency existed between them. In Finland, the class teacher was responsible for supporting students’ English, with help available from the English teacher only after problems emerged. In Korea, the English teacher was assigned to support the process. At the beginning of the project, students were encouraged to use English, but very quickly it was clear that they were unable to cope with the language barrier by themselves, even with teacher support. One teacher reported: “Language resources should have been thought out completely differently, our children do not yet have such strong language skills.” While Korean teachers thought language was the major barrier, Finnish teachers did not consider it a key difficulty in the project, as the primary objective was collaborative construction of knowledge and interaction in waste recycling. But the Finns also realized that English was not only pivotal in sustaining instructional activity, but also student motivation. Language support such as translation tools are needed for multilingual communities comprised of ten-year-old students (cf. Walker & Creanor, 2011).
Visualization of text (visual language) significantly facilitated construction of knowledge and interaction in waste recycling and reuse. Visual language enabled communication in the learning platform without words as it was built on students’ familiarity with sustainable development and recycling. Visual language could be employed in the construction of new knowledge.

Technical barriers

The learning platform Edu2.0 was new to all participants. Additionally, Finnish teachers had no previous experience of working in a cyber space. Support for technologic facilities such as internet connections and digital cameras is necessary, but not enough. A Finnish teacher mentioned that “The children brought their own digital cameras to school, but not everyone had cables for transferring photos.” Posting images in the Edu2.0 learning platform and defining image size was challenging to students. Technological infrastructure such as internet access, computing lab availability or laptops in classrooms should be the basic requirement in implementing this type of project. The best solution would be to provide mobile technology to students so that classroom activities are linked seamlessly to the project without having to go to the computing lab (cf. O’Neill, 2006). Convergent with activity theory (see Jones & Healing, 2010), the project demonstrated that individual players, teachers and learners are part of an activity system which should be thoughtfully integrated into a project system. The school system, learning environment, access to technology and infrastructure, and pedagogic solutions in learning projects all impact digital natives’ use of technology (Jones & Healing, 2010).

One Finnish teacher observed fourth graders’ (10-year-olds) need for technical support in the project as follows: “I often thought, hey c’mon, our IT Finnish society, some country is going to really pass us very fast, even though we think that our students are on the internet the whole time and can do whatever…they can only play on the net, what they know how to do is very limited, they don’t know how to save or search on the machine…. It is interesting to note that Korean teachers made similar comments. Successful infrastructure for internet connectivity, but poor application was mentioned by the Korean Presidential committee’s 2010 national IT agenda (Lee, 2010). In this project there was no homogenous net generation so to speak (Jones & Healing, 2010). Some students were able to assist their friends (peers) and teacher in the use of the new technology, but there were also students who needed support in digital technology use.

According to the Finnish teachers, students’ IT skills were limited, restricted to playing games on the internet (OECD, 2005). However, teachers may not fully appreciate how children live, act and experience the media world (Pohjola & Johnson, 2009). After being in the Edu2.0 platform with children, the Finnish IT support observed that boys quickly found more complicated applications and had fun with some functions commonly used in games. Games were familiar to students but ignored by teachers. Game-like features should be considered a valuable asset in learning.

Opportunities to employ IT facilities differ noticeably between Finnish schools. Reports indicate that there is a wide difference amongst institutions in device capacity, and clear regional differences also exist. Increasing educational use of ICT requires systematic support for facilities in school and structural and pedagogic development for teachers. Also, adequate internet connections at school and at home have to be guaranteed for all students (Niemi & Kumpulainen, 2008). Korean schools provided free H/W and network (internet access) to less affluent students as part of the social welfare system for educational equity. Most students, however, have internet access at home.

Collaboration barriers and cultural differences

In this pilot study, there was little interaction between Korean and Finnish students as the few resources were focused on producing products for the discussion forum. It was evident that students were not so interested in the visual products of others, but more interested in receiving responses to their own messages and work. Energy and resources were all used in creating one’s own products, which left neither time nor energy to comment on the work of others, although students loved reading comments on their work made by students from another country. It raises the need to emphasize commenting on other students’ postings as vigorous interaction.
Finnish and Korean students showed differences in communication style. These examples of international dialogue between school children show how Korean students use interactive dialogue to set challenges: “Hi! I saw Finnish friends’ mail. So I am very happy. I only have Korean friends but now I have some Finnish friends. Let's talk about our world.” Do you have a good idea about cleaning the world? I will not throw away the trash in the street. I am just 11 years old. So I think we cannot do difficult works but we can do very simple things in our homes and school. Saving water & saving Energy is the best way to clean the world.” The Korean student “throws out” a question for reflection. What can we do together to save the world? All in all, there were more questions on environmental issues in the Korean students’ messages, while the Finnish students’ messages were on average perhaps more descriptive; the latter focused more on presenting knowledge than challenging to dialogue on an emotive and attitudinal level (Munro, 2009).

This type of dialogue, rich context vs. low context communication, also shows that Korean students tried to receive consent from Finnish or other Korean students. Korean students tried to seek consent from other students in order to be socially accepted. According to a previous study on online interaction cultural differences among Finnish, USA and Korean university students, Kim and Bonk (2002) reported similar findings, though this was a study into cross-cultural differences in online collaboration behaviors in university students. In the study, Korean students showed a higher level of social interaction behaviors than Finnish or American students, who displayed very few social interaction behaviors. These differences are most likely related to different communication styles across cultures. It is generally understood that Western cultures, such as the U.S. and Finland, use low-context communication, and Asian cultures such as Korea use high-context communication. It is assumed that Korean students tried to use social context in order to make meaningful communication. Further research is recommended to ascertain whether this different style of communication is due to cultural differences or different teaching styles. Whatever the research outcomes, being sensitive to pedagogic cultural differences is essential (EBS, 2008).

Strengths of the operational model

Environmental education themes facilitate active learning

Environmental education and observations on recycling as a sub-section are easily applicable to authentic and problem-based learning. The learning tasks in this pilot were desirable for the following reasons: (1) they had a real-world relevance, (2) students had to investigate a single complex problem (tasks were not defined in detail beforehand, but were ill-structured, and not fixed in advance), (3) the learning tasks required a sustained period of time for investigation and could be integrated across subject areas (Herrington, Reeves & Oliver, 2010). In the operational model students worked in groups on open, real-world tasks and were (partly) able to play an active role in defining and solving learning tasks. Teachers had a guidance role in learning as defined by social constructivism. Students had an opportunity to network and learn through working collaboratively with students from another country (Kankaaranta & Puhakka, 2008; Herrington et al., 2010). In fact, students not only enthusiastically investigated the subject matter, but also culture. The project introduced the customs and culture of another school community. The pedagogic operational model facilitated crossing of traditional boundaries (cf. Hakkarainen et al., 2004) and took students outside their school—both near and far.

Knowledge visualization with technology

Technology can help us see things in a new way. Digital cameras and mobile phones make visualizing knowledge easy. Students can create documentaries which contain real-world phenomena in more visual form, giving them a better understanding of their own society (Jonassen et al., 2008). In this regard, knowledge visualization was successfully implemented in this project. A Finnish teacher narrates in the project’s early phase: “I have designed tasks for pairs in which they find out how recycling happens in their city. Students will interview the caretaker and catering staff on recycling and waste disposal at their school. Some will research Ekoroski’s website, some will produce information about Finland in pictures and words.” At the project’s conclusion one teacher felt: “An environmental theme was fantastic, and the images alone visualized some kind of interaction.” Traditionally, text-based virtual communication has been more natural for Finns (Munro, 2009). One key insight the project provided to teachers was the affordances of a virtual learning environment for visualizing knowledge. Images can strengthen authenticity and provide a real-world perspective. Positive learning outcomes are reported more often from learning
environments where visual technology, such as digital video technology and multimedia is applied (Vähähyppä, 2010). One characteristic of digital natives is that they are more oriented to visual media than previous generations.

Virtual learning environment

Teachers valued Edu 2.0 highly for international collaboration between elementary schools. The learning environment also contained elements students found interesting (gifts, profile images). All in all, Edu2.0 was a eureka-experience for Finnish teachers and students. A teacher’s experience: “It’s possible...to do whatever with the students there; they were very excited by the tasks we (teacher pair) had designed.” Another teacher’s summary of experiences: “Absolutely wonderful how the environment supports learning, starting with language learning.... The environment will continue to be used in classwork and in inter-class work.” Administration of the Edu2.0 platform worked extremely well, as students in the different countries were able to log in to the same place.

A teacher describes how enthusiastic students were to use Edu2.0: “... so the children paid attention to the platform, were extremely enthusiastic and inquisitive and even rather demanding...found out for themselves and supported their teacher.” A teacher explains: “When I showed one student how to attach a video, the information spread throughout the class.” The behavior of digital native school children is evident here; knowledge is received from peers and it is shared with peers. The role of learner and teacher alternates as necessary. Today, a teacher finds herself in the role of learner just as often as a learner finds herself in the role of teacher (cf. Hartnell-Young, 2006.)

Digital natives unreservedly adopt new technology and they have the courage to throw themselves into learning something new: “…what I have learned from students is to jump right in (teacher).” In Finland the ICT in a school’s day to day activities (“Arjen,” 2009) report recommends ICT be used more rigorously in teaching and learning. Children learn to utilize technology quickly, a natural characteristic of digital natives which teachers often tend to forget (Vähähyppä, 2010).

Motivation

Consistent with earlier studies (Korkala, 2009), students showed great motivation in using technology for environmental themes. The children were innately enthusiastic: “…after they had read the messages and images sent by the Korean children, it was hey, when can we go again...it was thoroughly interesting...” A meaningful and creative pedagogic use of IT engaged students in learning. The use of images was especially motivational. Finnish teachers felt the activities promoted learning: “The students said they had learned a great deal about recycling.” However, infrequent interaction dampened students’ enthusiasm. Thus frequent interaction is essential and teacher support can positively impact student participation. Sustained active teacher participation relies on intrinsic motivation which requires autonomy, competence, and relatedness (Radford, 2010). Korean teachers and students did not participate as enthusiastically as expected. This was attributed to a lack of language competence, as teachers and students found recycling related to their daily lives and teachers have full autonomy in class activities in Korea. Support for technology and English must be provided to teachers to ensure competence as suggested by the Self Determination Theory (Pink, 2009) and Radford’s (2010) motivation theory.

Teacher skills and attitudes have a critical role in how technology is utilized in teaching (Niemi & Kumpulainen, 2008). The three Finnish classes left without a partner class adapted the project to their own needs and felt the project was the best thing they did during the fall, even though it was not implemented according to its original objectives. Only one Finnish class did not complete the environmental education tasks or work in Edu2.0 (Pohjola & Johnson, 2009).

Students could acquire IT skills through authentic class activities, rather than formal instruction of IT skills. When content is of primary importance, the learning of which is supported by technology, authentic learning projects are created. A significant dimension of motivation was the articulation of knowledge, that the project helped students to develop writing skills for real audiences (Herrington et al., 2000; Jonassen et al., 2008). The Finnish teachers noted that even boys paid greater attention to grammatically correct writing when they knew they were writing for a wider audience in Edu2.0, contradicting the finding that female university students are better than male university students in cyber space asynchronous communication (Im & Lee, 2008). This invites further study on how school level and gender affect communication skills in a cyber space.
Pedagogy

Despite the potential power of technology, it is generally not utilized effectively in teaching. A primary reason for this is that teachers do not know how technology can be employed in meaningful ways. While teachers have the necessary technical IT skills, pedagogic training is also needed (Niemi & Kumpulainen, 2008; Vähähyyppä, 2010). This study showed that suitable themes and teacher training can motivate use of IT to enhance students’ authentic learning experience.

New learning concepts which emphasize a social construction of knowledge and collaboration demand changes in basic education in Finland. Technology has an important role in this reform process (Niemi & Kumpulainen, 2008). Pedagogic wellbeing is improved when teaching and studying are built on a participatory, listening, interactive and motivating learning environment that utilizes learning methods typical to today’s youth (“Arjen”. 2009). Changes in society, such as multiculturalism and the needs of personalized learning also set new challenges for schools. Kankaanranta and Puhakka (2008) suggest Finnish teachers use the advantages of networking and expanded learning environments afforded by IT specifically to acquire classroom connections to experts, friends and other educational groups. Korean educational policy addressed a similar need in education where creativity and personal character building education are emphasized.

It appears that the permanent impact of this project is the positive attitude of teachers towards international collaboration as a new learning environment and teaching method using IT (Korkala, 2009; Vähähyyppä, 2010). The positive effect of IT in education is evident when instructional activity in a learning cyber space is guided by clear pedagogic goals originating from genuine pedagogic needs (Vähähyyppä, 2010).

Operational model for international cooperation using cyber space

The pilot experience is applied to describe the operational model for international elementary environmental education in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Operational model for international elementary environmental education</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage</strong></td>
</tr>
<tr>
<td><strong>Activities</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

253
Reflection

The examined project is an example of a new educational practice in elementary school level international environmental education. The pilot study offers insights on how the use of technology can be strengthened in meaningful ways and identifies pedagogic models that integrate teaching to authentic contexts and promote authentic learning, as this is a necessary step on the road to such pedagogy (Niemi & Kumpulainen, 2008; Herrington et al., 2010).

ICT is a central part of life for today’s young people, the so-called net generation. Virtual interaction is a natural activity for digital native school children, but the challenge for project coordinators and teachers is to construct clear work methods for group-based collaboration. Construction of project implementation cannot bypass teacher-level reciprocal collaboration. The crucial significance of laying a solid foundation for collaboration also emerged strongly. In Finnish schools, limited ICT resources for the use of the entire class was a big barrier, while in Korea little support for teachers in organizational aspects due to lack of school leadership was a limiting factor. Korean teachers felt insufficient English skills to be the greatest obstacle.

New generations grow into digitalism and at school, young generations challenge the abilities and expertise of teachers (Vähähyppä, 2010). School children as digital natives adopt familiar functions from games in virtual interaction. While a teacher felt that students “only know how to play”, games actually provided students a context for perceiving the learning environment easily. New technology can help to put contemporary learning concepts into practice.

The attitudes of Finnish teachers and students towards the project idea, despite implementation difficulties, were positive. The project offered school children a virtual opportunity to become familiar with a significantly different culture and promoted self-awareness of their own identity (O’Neill, 2006). An attitudinal readiness for a new type of work for the 21st century, consisting of cognitive skills, working together, digital literacy, and communication skills (Cisco, 2010) exists. However, consistent with McNaught, Lam and Lam (2009), differences in students’ levels of motivation to use technology in learning were notable.

The pilot study, contrary to our expectations, found student activities more task-centered than interaction-centered. In future, thoughtful operational models to promote international interaction are needed. Trust building at the beginning of activities can enhance communication quantitatively and qualitatively (Im & Lee, 2007).

It would be interesting to compare the impact of culture on virtual interaction between school children. Korean students who are more rich-context communicators in general prefer communication through video or other forms of multimedia. Consistent with previous studies, the project examined here indicates that a western method of processing knowledge is traditionally more text-based, while an eastern approach relies more on knowledge visualization (Munro, 2009). This is consistent with the general notion that Asians (including Koreans) are rich-context communicators preferring to communicate in multimedia form. Communication differences need to be
culturally embedded in pedagogy, particularly in cyber space communication. French, Lee and Pidada (2006) have compared friendships of Indonesian, South Korean and U.S. youth in order to learn about group dynamics. In their study, Korean college students preferred small groups of 2 or 3, while college students from the U.S. preferred slightly bigger groups (5–6), and Indonesian students the biggest groups (8–10). Further studies should include how cultural preference for group size affects effective interaction in virtual learning environments.

Acknowledgements

This work was supported by the research grant of the Chungbuk National University in 2010.

References


Designing Hypercontextualized Games: A Case Study with LieksaMyst

Carolina Islas Sedano, Erkki Sutinen, Mikko Vinni and Teemu H. Laine
University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland // carolina.islas@uef.fi // erkki.sutinen@uef.fi // mikko.vinni@uef.fi // teemu.laine@uef.fi

(Submitted October 4, 2010; Revised February 10, 2011; Accepted April 10, 2011)

ABSTRACT

Digital technology empowers one to access vast amounts of on-line data. From a learning perspective, however, it is difficult to access meaningful on-site information within a given context. The Hypercontextualized Game (HCG) design model interweaves on-site resources, translated as content, and the digital game. As a local game design process, HCG builds on the resources available on-site (context) and transfers them beyond their origin (hyper). A successful example is the HCG stories in LieksaMyst, an application developed for the Pielinen Museum, in which the player is guided through a series of activities by a virtual host from the historic past. It took three years of extensive work and research for the LieksaMyst HCG stories to mature from brainstorming concepts to a fully-fledged museum service. Curators, educational technology experts and a total of 476 visitors (from 6 to 77 years old) contributed to the design process. The analysis of the context and feedback from the visitors enabled us to choose media, content and activities suitable to the Pielinen Museum. Our findings indicate that quality time, sincere reflection and communication between local experts and potential players, are indispensable when designing a HCG based game. The analysis benefits researchers and practitioners who are interested in the ways in which a game can bridge the gap between people and relevant on-site information.

Keywords

Hypercontextualized games, Storytelling, Museum, Mobile, Context

Introduction

“What fun... The program is humorous and encouraging!”
Female, 73 years, Finland

“It was a good simulation and I felt as if I was actually involved in the situation”
Male, 22 years, Nepal

“I have discovered history and a little more about the life of countryside people”
Female, 21 years, Russia

Above remarks were made by visitors to the Pielinen Museum. Currently the museum supports its visitors’ learning experience (independently of their age, gender and background) through different stories developed as a Hypercontextualized Game (HCG) named LieksaMyst. In essence, LieksaMyst is a digital application that provides immediate on-site extra visual information of selected items in the museum and guides the visitors’ understanding of living conditions in historical Finland. LieksaMyst is accessed and navigated by mobile phones. The HCG stories are the culmination of three years of research and development. In each story, a virtual character that lived in the same historical period as the house in which the game is played, hosts the visitor. In the course of the game, the virtual hosts share their daily life with the players and request the players to identify objects they need to execute their daily activities. These objects are to be found in the physical environment where the player is embedded. The ideal governing the design of the HCG stories in LieksaMyst is that players connect to and identify with the history that is enclosed in each one of the museum’s buildings.

The design process of the LieksaMyst HCG stories contributes to the research and development community in that it promotes an understanding of real life artifacts on-site and also facilitates access to meaningful information while the game is being played. Digital technology is considered a powerful tool, although it is obviously not the only relevant tool on-site. Our contribution offers a solution to an increasingly significant problem in our era, namely the lack of meaningful information on-site (Clark, 2010).

Digital technology is unquestionably changing the media landscape (Jenkins, 2006). By using a mobile phone with an Internet connection, one can tap into facts and user generated information from every possible location on our planet (Clark, 2010). Nevertheless, research shows that information overload overwhelms individuals (Schwartz, 2005) and the paradox lies therein that although one has access to considerable amounts of data, one can miss the transmission of deeper meanings. Thus, we require models that help us to integrate the different resources available
on-site to finally achieve a meaningful transmission of information to individuals. The Hypercontextualized Game (HCG) design model (Islas Sedano et al., 2010) aims, since its very conception, to link small audiences with their continuously changing context. In contrast, today’s digital games are developed outside the player’s context and their main aim is to target worldwide audiences. The game can thus be played anywhere and at anytime in a virtually generated world without a meaningful connection to the player’s reality.

We conceptualized LieksaMyst as a means to deliver meaningful information to the Pielinen Museum’s visitors on-site within the constraints of the available resources. Although museums often use modern technology to help their visitors understand their exhibitions, digital games are rarely used. Some of the existing game solutions for museums feature initiatives like the use of the technology and user acceptance (Ferris et al., 2004; Klopfer, Perry, Squire, Jan, & Steinkuehler, 2005). Complementary to this genre of museum games, this paper addresses a related but different problem. The question is how to design a game for a particular museum within the limits of its own available resources. Unlike generic technology, of which context-aware recorders that automatically explain the surroundings to the visitor are an example, we wanted to craft a game that originates from the particular context itself. Whereas conventional solutions are technology-driven, ours is driven by the context. The distinguishing perspective of this paper is thus the design process that we call hypercontextualized. A typical study executed on a technology-driven game is quantitative. It analyzes to which extent the designed solution, which is provided, promotes user experience or learning. In our case, the context is given, while the solution is designed alongside with and informed by the study. Therefore, our study is qualitative and explores the design process itself.

This paper identifies the elements that characterize LieksaMyst as an HCG and analyzes the design process. Each one of the HCG stories is designed and developed in, with and for the specific context. The stories target a small and inclusive audience, namely visitors to the Pielinen Museum. The LieksaMyst design process exemplifies a modern digital game carefully knitted with the reality of a specific context to deliver meaningful information on-site which, in turn, supports the co-creation of knowledge by its players. Furthermore, the design process highlights the players’ imagination and curiosity about the real life elements (e.g. environment, social dynamics) and utilizes these to guide the players in a positive learning experience. The carefully executed design process of the LieksaMyst HCG stories resulted in an enjoyable mobile service offered by the Pielinen Museum which the curators can expand upon as and when they wish.

Literature review

The combination of games and museums is nothing new (Hooper-Greenhill & Moussouri, 2000). Games are utilized without using technology widely in museums. Examples of this are the role-playing games at the Herren-Chimsee New Palace and King Ludwig II Museum in Bavaria and the treasure hunting games at the Van Gogh Museum in Amsterdam. Literature widely addresses the use and development of digital games in museums. However, most of these games are reported as temporal research studies and not as established services of the museums (Ferris et al., 2004). One potential reason for the limited presence of digital games in museums is the restricted involvement of curators in the design and development of the games. This leads to the infrequent transfer of research game prototypes to museum services. Additionally, researchers seem to want to produce edutainment by combining education and entertainment (Resnick, 2006). Instead, we argue, digital games should transfer and relay the curators’ passion and knowledge of the exposition to the visitors through play, as is the case of non-digital games.

According to Ravenscroft and McAlistier (2006), the enthusiastic application of digital games in education is apparently encouraged by the video game industry. Gee (2003), for example, argues that the knowledge gained by playing video games is transferable to other domains. In opposition to this claim, Ravenscroft and McAlistier explicitly state that today’s digital games are weak in linking the “game-playing activity to transferable social or conceptual processes and skills that constitute, or are related to learning” (2006). Additionally, knowledge cannot be viewed as a self-contained unit of information to be transferred without a contextual reference. Brown, Collins and Duguid (1989) hold forth that knowledge and cognition cannot be separated from a situation – instead, the activity and the situation co-create the knowledge that is developed and deployed at the moment in which it is to be learned. Cronje sustains their argument noting that “learning is constructed from the experience of the learner” (2006). Additionally, Robinson mentions that “we think about the world in all the ways we experience it, we think visually, we think in sound, we think kinesthetically, we think in abstract terms, we think in movement” (2006). Hence, to co-create knowledge alongside information we should involve ourselves in a situation – preferably at a multidimensional level.
The main challenge in any learning activity, including games, is to retain the learners’ engagement in the activity until they succeed (D’Mello, Chipman, & Graesser, 2007). The learner’s affective engagement in activities can give rise to a sense of belonging, acceptance and identification with something or someone (Sharan & Tan, 2008).

Hypercontextualized games

An individual resides in her or his own real life but can also decide, at any given time, to access the play-space and play a game while in this space. A game consists of game components, mechanics and dynamics. A game component is any unit essential to the functioning of the game. The game mechanics consist of game states and rules that define the allowed transitions. The game dynamics refer to all the possible ways in which the game mechanics are actualized.

A Hypercontextualized Game (HCG) is a locally designed game system (Islas Sedano et al., 2010) which helps its players to gain information about different subjects by using specific elements of the continuously changing context in the game. The HCG design model explains how each game is the result of interweaving three different perspectives: specific context elements (SCEs), subject matter information (SMI) and game system (GS). Each of these perspectives is described below. In this paper, we use HCG to refer to the design approach as well as games designed by following the said approach.

- **Game System** (GS) refers to the conceptualization of the core idea of the game. It takes advantage of the real life atmosphere to support the immersion of the player in the game. GS focuses on supporting affective engagement on-site.
- **Subject Matter Information** (SMI) determines the meaningful activities that the players can perform to co-create knowledge according to a clear aim and the resources available. Therefore, SMI focuses on the intellectual and pedagogical elements on-site.
- **Specific Context Elements** (SCEs) identify the potential game components in the specific context for which the game is designed and in which it is being played. This perspective classifies the components according to the environmental aspect of context (physical environment observable with our senses or instruments) and the intersubjective aspects of context (circumstances that are related to the communication of an individual with other individuals or with systems). Additionally, SCEs take into account the subjective aspects of context like affections, thoughts, ideas, reflections and meanings.

The HCG stories of LieksaMyst exemplify the previous definitions:

- **Game system** stimulates the visitors’ imagination and transports them into the past giving life to the exposition and involving them in the stories contained in each house.
- **Subject Matter Information** promotes the visitors’ reflection on historic Finnish lifestyles and adds to their understanding of the historical value of the on-site objects.
- **Specific Context Element** makes use of the objects normally displayed in the museum’s exhibition while respecting the museum’s atmosphere and rules.

The term *hyperpercontextualized* derives from the awareness that each game not only involves on-site contextual elements (SCEs) in its nature (GS and SMI), but also presents a context tolerance and context depth. *Context tolerance* refers to the flexibility of on-site contextual aspects that are in constant flux. The weather conditions or social dynamics that bear down on the museum are examples of this. *Context depth* is the degree in which a game involves the players in their context. Thus the term is a reminder of two aspects inherent to the games designed by it. Firstly, the games are based on and aimed at a particular context, which may be very limited. Secondly, the prefix hyper (from Greek υπερ, beyond) refers to the fact that HCG games are also tolerant in that they allow for applications in situations other than the original one. The games may also be played by different players from those who originally co-created them. The games thus move beyond the borders of the original conceptualization.

Digital games

Digital game titles are constantly increasing in the different genres that include adventure, fantasy and educational. We can identify three types of digital game designs for four platforms (Table 1).
Table 1. Digital game designers and platforms

<table>
<thead>
<tr>
<th>Designer</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game designer</td>
<td>Video games are located in the digital device and target global and mass audiences to offset their high production costs. With the development of the technology, some recent games can anchor virtual characters to certain real life aspects of the players. Location-based games make use of the technical capabilities inherent in mobile technology such as GPS, Bluetooth or WLAN. Most of these games have a research agenda, with the aim of connecting their findings to global products.</td>
</tr>
<tr>
<td>Puppet master</td>
<td>Alternate Reality Games (ARG) blur the distinction between real life and play-space through a strong and complex narrative that utilizes all possible media (digital and non-digital). ARG are mainly used promotionally and therefore temporally target wide audiences (Szulborski, 2005).</td>
</tr>
<tr>
<td>Game master</td>
<td>Pervasive Games are a combination of location-based games and ARG, with the clear distinction that these games “call” the player anytime the game wishes (Montola, Waern, &amp; Nieuwdorp, 2006). Up to now these games have been mainly research driven.</td>
</tr>
</tbody>
</table>

The HCG design model covers a previously unattended niche, namely games designed for small and focused audiences with the aim of delivering meaningful information on-site to their players during the course of the game. A HCG supports its players’ co-creation of knowledge by linking their real life experience with that of the play-space using the available resources. Table 2 summarizes the game design approaches, including HCG, for games that facilitate learning.

Table 2. Game design approaches by selection of characteristics

<table>
<thead>
<tr>
<th>Platform</th>
<th>Game designer focus</th>
<th>Puppet and Game master focus</th>
<th>HCG designer focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audience</td>
<td>Design of video and location-based games</td>
<td>Design for alternate reality games and pervasive games</td>
<td>Design for any platform</td>
</tr>
<tr>
<td>Availability</td>
<td>Up to millions globally</td>
<td>Research tests focus on groups, but the aim is to use the game for big to global audiences</td>
<td>Small and focused audiences</td>
</tr>
<tr>
<td>Production and sustainable costs</td>
<td>Most video games present high production costs</td>
<td>ARGs as well as most of the research interventions are expensive</td>
<td>Economically affordable and sustainable</td>
</tr>
<tr>
<td>SCE (Specific Context Elements)</td>
<td>Do not involve the player’s real life or on-site resources</td>
<td>Dependent on some of the player’s real life resources</td>
<td>Depends totally on the player’s real life on-site resources and context</td>
</tr>
<tr>
<td>GS (Game system)</td>
<td>Games do not involve the player’s real life on-site affective elements</td>
<td>Most games are fantasy role playing, creating a world that superposes the on-site reality of the player</td>
<td>The game makes use of the player’s real life resources that involve affective elements on-site</td>
</tr>
<tr>
<td>SMI (Subject Matter Information)</td>
<td>Games rely on virtual simulations</td>
<td>Hybrid use of resources (virtual and real life)</td>
<td>The game uses the player’s real life elements with intellectual and pedagogical relevance</td>
</tr>
</tbody>
</table>

Purpose of the study, limitations and research question

The HCG design model is the retrospective result of our research journey, and the model that guided us to design a game rooted in the specific context, namely the HCG stories in LieksaMyst. Therefore, the present article answers the question: What are the elements of the design process that make LieksaMyst a Hypercontextualized Game? The answer to this question benefits anyone with an interest in the development of digital games to support the players’ co-creation of knowledge on-site. Additionally, the design process of the HCG stories in LieksaMyst indicates that the transmission of meaningful information on-site is applicable to any educational field and that the game can be sustainable and affordable in any context as long as the game is rooted in it.
Most of the literature, related to digital games, focuses on issues like the use of (existing) technical solutions, their usability, game experience (mainly in commercial games) or the game’s teaching potential. Additionally, as we mentioned earlier, most of the digital game designs are aimed at mass audiences. These challenges call for quantitative studies.

The neglected part of the research has been, according to our knowledge, the design processes of successful digital games for small audiences that deliver meaningful on-site information that supports situated learning. Therefore, the purpose of our study has been to explore and design a digital game that is intentionally rooted in a specific context. We maintain that the real world facilitates the learners’ immersion into a specific subject during their museum visit. This task requires mainly a qualitative research approach.

On-site background

LieksaMyst was designed for the open-air section of the second largest open-air museum in Finland (Figure 1). The museum contains more than 70 buildings and structures arranged in an area covering almost two hectares and it is home to over 100,000 objects that were used in different historic time periods. These days Finland is viewed as a technological country, which boasts high levels of education and solid wealth creation post World War II. This relatively quick economic growth, as perceived by foreigners and experienced by nationals, is understandable when one visits the museum.

<table>
<thead>
<tr>
<th>Table 3. Pielinen Museum statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visitors during the year</td>
</tr>
<tr>
<td>Permanent staff</td>
</tr>
<tr>
<td>Temporal staff</td>
</tr>
</tbody>
</table>

In addition to the large outdoor section, the museum has a smaller indoor exhibition. The open-air exhibition is open mainly in the summer time (May 15th to September 15th) and closed for the rest of the year due to cold weather. In the summer months one hour long guided tours are offered as well as extra activities like a Children’s Day. Table 3 summarizes the museum’s visitor and staff statistics for the duration of the LieksaMyst design process.

Research design

The case study was performed using a social constructivism paradigm. We needed to understand the different views and backgrounds of the museum’s visitors (Saunders, Thornhill, & Lewis, 2009). We were aware that individual
worldviews are socio-cultural-historically negotiated and formed through interactions (Engeström, 1987). For this reason, the views and opinions of the participants in our study helped us to understand the phenomena under investigation. A case study was used because it is an empirical investigation done in real life conditions using multiple sources of data. The qualitative data (e.g., open questions from questionnaires and informal interviews) was analyzed using content analysis, and the close-ended type of questionnaire section was analyzed using a quantitative analysis. Mixed methods were used to support a complementary view of the study (Saunders et al., 2009).

Figure 2. Data collection supports two processes: understanding the needs of the Pielinen Museum and concretizing LieksaMyst as a product

Data was collected from 2007 to 2010 (Figure 2). Each year’s data collection supported two processes, namely the understanding of the different context aspects relevant to the Pielinen Museum and the actual design of the application.

Participants

The objective was to involve all the stakeholders, which included curators, visitors, and educational technology researchers, in a common aim of developing a new application for the museum. Every stakeholder’s contribution was documented and taken into account. Due to the diverse roles of the stakeholders, their participation in the decision process differed.

Visitors’ opinions and expectations are crucial to the application but their stay in the museum is temporal. Visitors also hail from different backgrounds, genders and nationalities. The sample of the 476 feedbacks gathered is described in Table 4.

Curators are museum and content experts who are knowledgeable and passionate about their museum. They have resource expertise, content knowledge and are preoccupied with the learning experience of their visitors. Their contribution is fundamental to the creation of a relevant and well-structured application.

Educational technology researchers supply the expertise and skills needed to bridge the gap between the curators, visitors and resources available (information, object and technology). In this case two of the experts were Finns (one native to the region) and the third Mexican.
Table 4. Visitor sample description

<table>
<thead>
<tr>
<th>Phase</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54 volunteer visitors between the ages of 9 and 74 years. 13 foreigners.</td>
</tr>
<tr>
<td>2</td>
<td>119 volunteer visitors between the ages of 6 and 77 years. 43 foreigners.</td>
</tr>
</tbody>
</table>
| 3     | Three test groups were organized:  
  a. A group of 17 students aged 13 years.  
  b. Seven seniors (between 58 and 73 years old) from the community.  
  c. A group of 15 foreigners aged between 19 and 44. |
| 4     | Informal tests:  
  Nine participants (4 foreigners, 5 locals) completed the phase 3 survey.  
  29 exchange students (11 and 12 year olds) played the game.  
  Focus interview with a family of five members. |
| 5     | 101 Finnish students aged 11 and 12 years old participated in an educational study.  
  69 volunteer visitors (ranging between 8 and 73 years old). |

Tools

The tools utilized in the HCG design process can be classified in two clusters: the data collection tools and an instrument developed to aid the creation of the HCG stories. Both will be briefly defined.

Data collection tools

Each phase had its own objective as regards the understanding of the context and the creation of the artifact. Table 5 briefly describes the methods used in the design process as per phase (I=Interviews; IF=focus group interviews, Q1=pre questionnaire, Q2=post questionnaire, D=informal and formal discussions, P=photos, OT=object tags).

Table 5. Data collection

<table>
<thead>
<tr>
<th>Phase</th>
<th>Methods</th>
<th>Visitor data collection</th>
</tr>
</thead>
</table>
| 1     | Q2      | Unattended surveys.  
  Survey topics: reasons for visiting the museum, rank museum’s interesting areas, perception of the exposition’s clarity. Additionally, sensing the play habits of the visitors and how would they improve the museum. |
| 2     | Q2-D-P-OT | The researchers carried out workshops with participatory design approaches (Schuler & Namioka, 1993) for volunteer visitors in five selected areas of the museum. Each location was scheduled for one day. The workshop was personalized and consisted of the following steps:  
  a. Visitors selected objects displayed in the location and wrote down what they would like to know about them. The anonymous cards were left on their selected objects.  
  b. Researchers administered a pre-made quiz on a mobile phone about facts pertaining to the location. The aim was to listen to the visitors’ opinions about the application and any other ideas.  
  c. Visitors filled in a survey which included the following topics: preference of media used, senses and activities, their opinion on other digital tools used in museums, their ideas for the museum’s improvement, their reasons for visiting the museum, and the information they would like to know in the specific location. |
| 3     | IF-Q1-Q2-D | The test groups filled in one pre and one post questionnaire.  
  Pre questionnaire topics: the visitors’ mobile phone usage and media preferences, their opinion about museums and digital tools in museums.  
  Post questionnaire topics: the visitors’ motivation to play the game, their most enjoyable and difficult game experiences, activities they liked and suggestions for improving the game. Additionally we evaluated different usability elements, visitors’ visions of the technology and what caught their attention during their visit. |
Phase | Methods | Visitor data collection
--- | --- | ---
4 | IF-Q1-Q2-D-P | Same questionnaire as in phase 3. The focus interviews aim to understand the game experience of the players.
5 | Q2-D-P | Survey given to the players after playing. Survey topics: visitor’s motives, self-evaluation of their learning experiences, their opinion about the game and the museum, and their suggestions for the improvement of both.

**Instrument for the creation of the HCG stories**

Based on the visitors’ requests after the first two phases (Figure 2), we decided to create HCG stories that would promote a conversation between the past and the present through the mobile phone and in which the visitors would actively take part in on-site activities. These activities are interwoven with the HCG stories.

Each activity is a sequence of individual actions (Engeström, Miettinen, & Punamäki, 1999). Thus, each HCG story contains the actions needed to complete an activity. For example, if a virtual host needs to make coffee, the actions involved in the activity are toasting coffee beans, grinding them, and then putting them in a pot in order to brew the coffee. As the series of actions progresses, the result of the activity is shown to the player. The player must also perform different actions in real life to keep the story going. These actions are, for example, finding the objects and entering their code into the mobile phone when requested. We call these actions *player interactions*. The player receives feedback once s/he has interacted with the system. The feedback matches the player’s activity as well as that of the HCG story.

Each HCG story in LieksaMyst is told through a combination of scenes. Every scene has a clear learning objective related to Finnish history and utilizes objects displayed on-site. Most of the activities, objects and media used are the ones that the visitors reported they preferred (*Table 5* and *Figure 5*). The scenes are strung together to offer a vivid lifestyle experience of a specific historic period.

![Figure 3. Instrument used for developing LieksaMyst stories](image)

The format of the communication instrument (*Figure 3*) between curators and researchers allowed us to share a common language. It permitted us to easily detect the programming and the storytelling structure and allowed us to identify the embedded on-site elements, the learning aim, and the media involved to stimulate different affections or intellectual challenges. For example:

1. **Scene** – specifies the number, title, and location of the current scene in real life. It also indicates the location, time and hour of the scene in the host’s life.
2. **Storyline** – describes the dialogue between characters in the story (e.g., virtual host). It includes references to complementary media used in the conversation such as photos and audio (special effects or music).

3. **Activity** – specifies the host’s activity in this scene, e.g., making coffee.

4. **Actions** – describes actions that sustain the host’s activity, the objects used in the activity and a brief explanation of the activity.

5. **End of the activity** – delivers the results of the host’s actions and the player’s actions in a visual format. If possible, it is complemented by a sound effect.

6. **Interaction** – indicates the type of interaction that the player has in the game system, for example finding the toaster or the grinder.

7. **Feedback after activity** – clarifies the type of feedback received in point 5 (end of the activity).

8. **Learning goal** – states explicitly what the learning goals of the scene are. For example, understanding the working of a floor loom and a “rissa” (Finnish word denoting part of the loom).

9. **Follow up** – indicates the following scene(s).

### Procedure

We used more than one mode of data collection (Table 5). All data was captured digitally and we followed a **qualitative interpretational** approach (Neuendorf, 2002). To establish trustworthiness, two techniques were utilized, **prolonged engagement** and the **triangulation of methods** (Lincoln & Guba, 1985).

### Design: From brainstorming to LiekkaMyst

#### HCG stories in LiekkaMyst

The HCG stories in LiekkaMyst are based on historical fictional narratives or fictitious stories that are set in the past. These narratives are supported by historical circumstances and detailed established facts. Thus, we can use the existing physical historical atmosphere of each story to stimulate the imagination of the visitor (now player) and so we bring life to the exposition. Different virtual characters from the past act as the player’s host. In this sense, it is a type of role-playing game because the players take the role of the guests who listen to and help their hosts.

The virtual characters and players communicate by means of a mobile phone (Figure 4). The game’s enchantment starts when the players, in the present, have a conversation and help their virtual hosts to perform their daily activities in the past. The players support their hosts by locating different objects needed to complete their tasks. For example, Jussi (a forestry worker at Pusurinjoki camp), asks the player to find his work tools. Hence the virtual hosts in the HCG stories converse with the player and, in so doing, they explain their life style in the past. The border between past and present is blurred when the players seek and identify objects to help their hosts.

It is relevant to mention that “players” do not refer to a specific group of children or adults. Instead, we promote inclusion by welcoming anyone who wants to play and who has the skills required to operate a mobile phone.

#### Phases of the design process

*Figure 5* sketches the five development phases of the HCG stories. Its foundation is the analysis of the information gathered in the first two phases, which focused mainly on the understanding of the different context aspects (e.g., environment, user preference of media, types of information). Our design decisions were sustained by voluntary visitor contributions and influenced by those activities that the visitors could do on-site (in accordance with museum regulations) supported with the available media (local resources).

Once the first prototype was ready, we tested it in Phase 3 with three different groups (*Table 4*). The prototype was unanimously well accepted by the testers, and it was workable within the Pielinen Museum. Phase 4 focused mainly on building the technical infrastructure, completing the second story and refining both stories so they could be used in LiekkaMyst. Finally Phase 5 focused on installing the latest version of the LiekkaMyst and offering it as an official museum service. The process included incorporating the museum infrastructure and training museum personnel to
operate the game without the help of the researchers. A story editor was developed in Phase 4 and 5 with which the curators can continue to create stories for the museum. In the first summer (2010) that LieksaMyst ran as an official service of the museum, a preliminary analysis of data showed that most visitors demanded more stories and even made suggestions as to which type of themes should be incorporated.

![Diagram](image)

**Figure 4. Screenshots of the HCG stories in LieksaMyst**

**Technical perspective**

In the first two phases, we took decisions regarding the digital technology in response to the circumstances, for example, the decision to use mobile phones and the Myst platform (Laine, Vinni, Islas Sedano, & Joy, 2010). The Myst platform is based on a client-server approach, in which the content is stored and updated on the server side and pushed to the clients’ devices over the network in xml format each time they play. We decided to use “old” fashioned interactions due to context characteristics. For example, today’s RFID technology requires close contact between phones and objects in order to trigger an action. This would violate the museum’s “no touch” policy, and phone models capable of reading RFID tags are currently rare. Therefore, we designed wooden tags to support the interaction between the player and the environment. The solution paid heed to the museum rule and was effective in that it blended in with the environs.
Table: Summary of the five phases' observations.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Aim per process</th>
<th>Participatory design</th>
<th>Highlights per process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1, 2007</td>
<td>a.1 General understanding of the context aspects</td>
<td>- Perceived the museum to be easy to understand.</td>
<td>- Museum's area is wide, thus visitors selected their major interest locations.</td>
</tr>
<tr>
<td></td>
<td>b.1 To gain resources for transforming what type of applications is suitable.</td>
<td>- Requested more information related to activities and objects.</td>
<td>- The museum's exhibition does not threaten the visitors; on the contrary, they enjoy it, but they requested for support to understand it.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Requested outcome, and to bring the piles alive.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Proposed to keep the environment as it is, visitors enjoy the pace of the place.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Proposed that the information is given on computers; they will not read it.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Proposed to do not play games, if they do not want to.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Suggested to add video, animation, more interaction and personal control.</td>
<td></td>
</tr>
<tr>
<td>Phase 2, 2006</td>
<td>a.2 To gain understanding of the context aspects in the selected areas</td>
<td>- Reason to be in the museum: 62.18% of the visitors due to historical interest, and 29.41% of them due to curiosity.</td>
<td>- Visitors supported the idea to develop a mobile phone application for the museum.</td>
</tr>
<tr>
<td></td>
<td>b.2 To gain concrete facts that can be involved in the concept design and prototype</td>
<td>- Mean interest in human facts than in objects.</td>
<td>- Content should support a diversity of background, variety of personal interests, and individual pace of participation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 57% reported to like using imagination mostly by stories, putting hands on and images.</td>
<td>- Museum technical infrastructure is poor and challenging.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Choose media to receive information was photos, followed by text.</td>
<td>- Visitors want to know mainly about daily activities and the use of objects.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Selected to use their vision followed by their hearing to understand the exposition.</td>
<td>- Visitors tag the objects that call their attention with the following recommenations: What, Who, Where, Why, and How.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Interactions activities are: (1) matching with objects, (2) matching information and (3) having stories.</td>
<td>- Stories, working with objects, matching information, text, photos are the basic elements to estimate the understanding through the vision of the visitors.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Few visitors had experienced the use of digital tools in their visit, and their opinions about it are positive.</td>
<td>- Design of LieksaMyst concept and its prototype by curators and educational technology researchers based on data and experiences gathered in phases 1 and 2.</td>
</tr>
<tr>
<td>Phase 3, 2006</td>
<td>a.3 Integrate the tool with the context aspects</td>
<td>- Reported to feel involved in the stories: “the artifact was clearly integrated into the game despite its simplicity.”</td>
<td>- Integrates and intertwines the environmental context by using the surrounding and do not ignore it.</td>
</tr>
<tr>
<td></td>
<td>b.3 To gain concrete facts that can be involved in the concept design and prototype</td>
<td>- Regardless that the last was in winter, the house was sold, dark and emptier than in summer. Visitors described liveliness of the house: “I feel I talked with Anna and helped her.”</td>
<td>- Integrates and intertwines the inter-personal context by respecting the pace and space of different individuals.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acknowledgment of personal learning experience about the history in Finland.</td>
<td>- Personal learning experience about the history in Finland.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Recognition of the interactions in the house.</td>
<td>- Personal learning experience about the history in Finland.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Intention of a very traditional museum and activities and very matter technology.</td>
<td>- Personal learning experience about the history in Finland.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Request to include more languages and photos.</td>
<td>- Curiosity was the primary motivator that impelled the visitors to play the game until the end, followed by personal control.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Suggested it was easy to use the phone, no usability problems among all the age groups.</td>
<td>- Searching for objects was the most preferred activity.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- To adjust the difficulty level of the game and quality of media assets.</td>
<td>- To adjust the difficulty level of the game and quality of media assets.</td>
</tr>
<tr>
<td>Phase 4, 2006</td>
<td>a.4 To install the museum technical infrastructure needed for LieksaMyst</td>
<td>- From diverse unofficial tasks and presentations we gathered informal feedback (45 individuals).</td>
<td>- To install the technical infrastructure took a considerable amount of time.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Editor for the creation of more stories started to be developed.</td>
<td>- Editor for the creation of more stories started to be developed.</td>
</tr>
<tr>
<td>Phase 5, 2010</td>
<td>a.5 To transfer LieksaMyst complete in hands of the museum</td>
<td>- Visitors who played had not complained about the game.</td>
<td>- To use the museum's infrastructure required some adaptation process.</td>
</tr>
<tr>
<td></td>
<td>b.5 Enable curators increase the HCG stories</td>
<td>- Players feedback does not focus on lack of массив of information, but they request more stories.</td>
<td>- Integrate LieksaMyst as a current service of the museum implies to train the temporal personnel.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- It is challenging to transmit to the visitors in the enhance of LieksaMyst.</td>
<td>- Museum started to created more stories.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Visitors suggest themes for the stories.</td>
<td>- Visitors suggest themes for the stories.</td>
</tr>
</tbody>
</table>

Figure 5. Summary of the five phases’ observations. Each phase includes the aims of (a) understanding the different context aspects at the Pielinen Museum and (b) designing and developing the LieksaMyst HCG stories.

Conclusions

The analysis of the LieksaMyst design process yielded the following recommendations for the stakeholders involved:

- **Museum staff**
  - Consider the feedback from visitors to improve museum exhibits and visitor experiences.
  - Incorporate interactive elements that align with the educational objectives of the museum.

- **Curators and Educational Technologists**
  - Collaborate more closely with museum staff to ensure that stories and interactive elements are relevant and engaging.
  - Continuously update and refine stories and interactive components based on visitor feedback and educational requirements.

- **Visitors**
  - Expectations should be aligned with the museum’s core values and educational goals.
  - Feedback should be actively sought and acted upon to enhance visitor experiences.

- **Advisors and External Partners**
  - Provide guidance on how to integrate educational technologies effectively.
  - Offer expertise in content development and user interaction design.

By implementing these recommendations, the museum and its collaborators can further enhance the educational and visitor experiences, ensuring that the LieksaMyst project remains relevant and engaging.
Visitors (end users) must be heard. Even when catering for a small audience, diversity of needs makes the task of transmitting tailored information challenging. Relevant patterns emerge while listening to the visitors. Researchers should be willing to listen and understand a wider scope of elements beyond their own agenda. In our case we had to take into account the economic circumstances of the museum, how it operates and the knowledge enclosed within its walls.

Curators’ openness and willingness to listen and collaborate with the educational technology researchers is fundamental.

Robinson writes that “original ideas that have value – more often than not come about through the interaction of different disciplinary ways of seeing things” (2006). The HCG design process is a concrete example of his statement. The achievement of the HCG stories to deliver meaningful information on-site is a result of understanding the context from different perspectives and finally harnessing all views to help support the players in their knowledge creation. Therefore, quality time, reflection and communication are the building blocks that lead to the formulation of the LieksaMyst stories.

1. *Quality time* is required to thoughtfully identify and analyze the different resources available on-site (SCEs) in harmony with the application’s purpose (SMI and GS). For example the HCG stories of LieksaMyst show the quality of time invested in the design process, even before brainstorming (Figure 2 and Figure 5).

2. *Reflection* is needed when using the contextual information to immerse the players in situated activities. The game should support intended emotions, experiences, and knowledge co-creation by using on-site resources. In our case study we constantly reflected on our decisions (Figure 5) to assure that the museum supports the players’ immersion in the HCG stories and stimulates the creation and assimilation of knowledge on-site.

3. *Communication* is fundamental if one wishes to capture the passion and knowledge of the creators in the game system (Islas Sedano et al., 2010). Consequently, an affective and emotional engagement is promoted between the authors and the players within the location in which the game is rooted. For example, currently LieksaMyst’s players are suggesting themes that can be incorporated in the stories. Curators, with the infrastructure installed, foresee different possibilities to transmit their passion: “LieksaMyst opens up interesting possibilities... now my imagination is running free as well!” (S. Harjunen, personal communication, August 5, 2010)

To illustrate the contribution of the HCG design model we evaluate two games as end-products with similar characteristics. We utilize the HCG’s perspectives as a basis for comparison (*Table 6*). The games are the HCG stories of LieksaMyst and REXplorer which is an “interactive mobile and cross-media cultural heritage adventure game for tourists” offered in Regensburg, Germany ("Media," n.d.).

<table>
<thead>
<tr>
<th>HCG perspectives</th>
<th>HCG stories in LieksaMyst</th>
<th>REXplorer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCE (selected context elements)</strong></td>
<td>Players use their personal or museum mobile phone to play the game. They will interact with specific on-site resources of the Pielinen Museum.</td>
<td>Once players have the game controller, they can play anywhere as the game does not demand interaction with specific on-site elements. Players can be standing next to a specific building while listening to information about it, or they can be somewhere else.</td>
</tr>
<tr>
<td><strong>SMI (subject matter information)</strong></td>
<td>The game delivers historical information regarding the Finnish lifestyle and the players should understand this content by performing situated activities to continue the game.</td>
<td>The game delivers information and facts to the player related to specific buildings, without any further activity to support the comprehension of the material. Players can generate personal location-tagged photos of any location that interests them.</td>
</tr>
<tr>
<td><strong>GS (game system)</strong></td>
<td>The game combines a historical narrative with activities using on-site resources and thus attempts to connect the player’s emotions to the past.</td>
<td>The players simulate magic spells, with the help of the controller, to wake up spirits to tell them stories. Therefore the players’ attention is distracted from the on-site real facts to the superimposed fantasy narrative and its supported elements, e.g., magic spells with the technology.</td>
</tr>
</tbody>
</table>
Table 6 shows that a HCG focuses on and makes use of the contextual elements in which the game is rooted. Technically REXplorer can easily increase its context-aware interactions. However, it would require time, reflection and communication with local experts and potential players in Regensburg to transform REXplorer from a digital game that only delivers “local” information into a HCG game. The process would start by asking which on-site elements can promote individuals’ immersion in the context and support their knowledge co-creation, instead of only using generic fantasy elements with localized information.

Although our research yielded positive results, there were some real limitations. Firstly, because of its position (526 km northeast of Helsinki), the museum draws significantly less visitors than a museum located in a metropolis. Consequently the study population is relatively small. Secondly, the researchers involved in the design process also conducted the assessment surveys and thus a certain degree of subjectivity is unavoidable. In general, the small number of users is not a problem for qualitative research, but it does limit the general reliability of the results. Therefore, we are working on the evaluation of other games utilizing the HCG model.

The HCG stories in LieksaMyst are now being offered as a museum service but the evaluation has just started. It is fundamental to understand the impact of LieksaMyst as a longitudinal study. Researchers need to differentiate between the experiences of visitors who play the game for fun or as part of a school/work project. Additionally, as technology evolves, we must reflect on how to keep LieksaMyst up-to-date. We need to employ new resources and media in order to include other types of visitors with different learning and game preferences, while LieksaMyst must be kept sustainable and affordable. Last but not least, it is essential to apply the HCG approach to other contexts and fields in order to gain an understanding of how to successfully design games rooted in the player’s context.

Acknowledgments

The authors of this paper would like to express their gratitude to the personnel of the Pielinen Museum and their visitors; especially to Liisa Eskelinen and Marketta Haavila who have supported us unconditionally. We also want to thank the National Board of Antiquities that has partially funded us.

References


Understanding Social Nature of an Online Community of Practice for Learning to Teach

I-Chun Tsai
Educational Foundations and Leadership, University of Akron, Buchtel Commons, Akron OH 44325 USA // ichuntsai6@gmail.com

(Submitted October 18, 2010; Accepted June 15, 2011)

ABSTRACT
This study is aimed to explore the social nature of membership in an online community of practice (NETwork, Nurturing Elementary Teachers’ work) whose purpose is to support pre-service and in-service teachers with a collaborative virtual space for learning how to teach. Path analysis was employed to explore the relationships among social constructs: perceived ease of use and usefulness, social ability, sense of community, satisfaction with NETwork experience, and effectiveness of NETwork for Teaching. The results show members’ satisfaction was the only direct factor impacting members’ perceptions of effectiveness; several social constructs have direct impact on members’ satisfaction; and technology acceptance influences members’ social ability, sense of community, and learning satisfaction.

Keywords
Teacher online learning community, Sense of community, Social ability, Technology acceptance, Online learning satisfaction

Introduction
The rapid development of the Internet has led to expansion and enhancement of online learning in higher education (Allen & Seaman, 2010). Recent years, researchers have been exploring alternative forms to traditional course-based learning. One such form that has potential in the field of teacher education is the use of online communities of practice (CoP). Researchers in teacher education have particular interest in CoP as a method for diminishing the gaps and disconnect between the stages of teachers’ professional development. Supporting a professional continuum of learning that spans pre-service teacher education, induction of beginning teachers, and continued professional development is a key challenge of teacher education (Feiman-Nemser, 2001). Since 1990s, teachers’ professional development has attempted to move beyond simply supporting teachers’ knowledge and skill acquisition by changing the training format from one-shot professional development to long-term professional development (Vescio, Ross, & Adams, 2006). The establishment of professional learning communities (PLCs) based upon a CoP framework (Wenger, 1998) has been recognized as an effective model to support the reform of teacher practice and teaching reflection (Hollins, McIntyre, DeBose, Hollins, & Towner, 2004; Strahan, 2003; Andrews & Lewis, 2002).

However, participating in the PLCs required teachers to travel to particular locations and coordinate tight schedules for meetings. Taking time to meet is problematic because teachers indicate that teachers’ daily teaching schedules occupy most of their time (Scribner, 2003). Improved ease of access and capabilities of network technologies offers the potential to support PLCs online where teachers can participate at anytime and anywhere. Some cases of applying online systems to support the continuum of teachers’ professional development have been found effective but not easy to sustain. (Tsai, Laffey, & Hanuscin, 2010; Job-Sluder & Barab, 2004; Gray & Tatar, 2004; Steele, 2002; Desimone, Porter, Garet, Suk Yoon, & Birman, 2002). These studies indicated difficulties to sustain members’ feelings of a sense of ownership, connections, trust, commitment to the community, and a sharing culture (Tsai, et al., 2010; Barab, Makinster, & Scheckler, 2004; Schlager & Fusco, 2004) which were crucial social factors impacting members’ interaction and retention (Carr, 2000; Chyung, 2001). To understand how to better sustain teachers’ interaction and participation in an online CoP, this study investigated the interdependent relationships among the social factors of the online learning community, NETwork.

Theoretical perspectives
Community of practice
Shaffer and Amundsen (1993) define the term: community as “a dynamic whole that emerges when a group of people participate in common practices; depend upon one another; make decisions together; identify themselves as
part of something larger than the sum of their individual relationships; and commit themselves for the long term to their own, one another’s, and the group’s well-being” (Shaffer & Amundsen, 1993, p. 10). A form of CoP that has a major focus on improving practice and supporting the learning of members is also called a learning community. To distinguish learning communities from other groups, Woodruff (1999) suggested four elements, including function, identity, discursive participation, and shared values, as primary elements that unify a community.

A form of CoPs that use the Internet as the primary mechanism for communication, participation, and sharing is called an online community. Preece (2000) described online communities as consisting of people, shared purpose, policies, and computer systems which diminish members’ concerns of location and time. Wenger (1998) defined CoP as groups of people who join together with a common purpose and share a common practice. Members of CoP integrate practice, meaning, identity, and community as components of learning and knowing in their interaction within CoP. Lave and Wenger (1991) said “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). Communities are places where practitioners can share interests and resources, engage in joint activities, and work toward the same goals. Individual’s growth depends not only on members’ individual inputs but also the shared values and knowledge reproduced in the interaction. Learning and membership in a community are intertwined as members’ identities change gradually from peripheral participation to more core roles as they gain more knowledge of practice. The changes in members identities, indicating how they perceive who they are and how other members think about them, is a result of their participation and engagement in online learning and interaction (Wenger, 1996). Members’ learning, identity changes, and relationships with others are influenced by their feelings and levels of participation and vice versa.

Social factors influencing interaction

When online PLCs operate through computer-mediated communication (CMC) tools and systems, how to facilitate and foster online social interaction within the CoP becomes a critical factor for the success of teachers’ professional development in the online learning community (Gess-Newsome, Blocher, Clark, Menasco, & Willis, 2003). Researchers found when people interact through online learning environments, the nature of the tools and the social constructs established around the contexts influence members’ participation and interaction (Picciano, 2002; Rovai, 2002a). The usages of CMC tools have shown to influence students’ participation and interaction in online learning (Collins & Zane, 1996). Similarly, Hara and Kling (2000) discovered that communication and technical difficulties experienced by students in online learning environments frustrated them and impeded interaction. This is critical because levels and types of interaction influence members’ sense of community (SOC) and participation (Lally & Barrett, 1999).

Previous studies reported that without sufficient social interaction, students experience a sense of isolation even though they do appreciate the flexibility and convenience of the virtual learning space (Rovai & Wighting, 2005; Cereijo, Young, & Wilhelm, 2001; Curry, 2000). Students’ feelings of isolation were found to be a primary reason for higher dropout rates in distance education (Carr, 2000). A lack of social interaction was a factor that depressed students’ satisfaction in online learning (Arbaugh, 2000). Students’ perceptions of their interaction and sense of presence were found to have positive relationships with their perceived performance in an online environment (Picciano, 2002). To be more specific, researchers in teacher education discovered that participants’ interaction and participation in CoPs were associated with the effectiveness of their learning, collaboration, and how they can apply what they learn (Job-Sluder & Barab, 2004; Gray & Tatar, 2004; Schlager & Fusco, 2004). The following sections identify previous studies that help form a model of how people’s SOC, social ability (SA), and technology acceptance are primary constructs impacting social interaction and participation of CoP.

Sense of Community (SOC). SOC and SA have been identified as two critical factors influencing members’ level of online participation and social interaction (Lin, Lin, Liu, Huan, Shen, & Laffey, 2006; Rovai, 2003; Picciano, 2002; Carroll, 2001). McMillan and Chavis (1986) defined SOC 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). SOC is one of the factors that can make an online community a learning community (Blanchard, 2000; Haythornthwaite, Kazmer, Robins, & Shoemaker, 2000). Prior research reported learners benefit from being a member of a community by feeling a sense of belonging and having others to ask for
support (Wellman & Gulia, 1999). Communities have greater flow of information among members, availability of supports, commitment to group goals, and higher members’ collaboration and satisfaction when people experience a stronger SOC (Scott, 2004; Wellman, 1999; Dede, 1996; Bruffee, 1993; Tinto, 1993). Similar to the results from studies of online courses, SOC has been shown to be a factor in sustaining social interaction in virtual teacher communities (Job-Sluder & Barab, 2004; Steele, 2002).

**Social Ability (SA).** SA is defined as “a person’s capacity to associate with fellows and to use the members, resources and tools of the social context to achieve something of value” (Social Computing Research Group (SCRG), 2006, p.2). Social presence (SP) and social navigation (SN) are underlying factors of social ability in online learning environments (Laffey, Lin, Lin, 2006; Yang, Tsai, Cho, Kim, & Laffey, 2006). In online learning studies, SP “as an attribute of computer-mediated activity is derived from media studies about how effectively media (TV, etc.) convey the sense that mediated participants were really present” (Laffey et al., 2006, p. 166). SN is defined as “a construct representing being aware of what others are doing as a primary guide for one’s own actions” (Laffey et al., 2006, p. 166). Prior research has found positive relationships between students’ perceived SP and assignment scores (Picciano, 2002). Additionally, what members do or how they act in an online environment are found to be based upon what others have done, where others are, and what they have looked at (Gutwin & Greenberg, 1998; Dourish & Bellotti, 1992).

**Technology Acceptance.** Social constructs of online learning are highly associated with the affordance of technology in the online learning environment. The CMC tools influence how members socialize with others and appropriate tools when interacting and participating online. With the intention of understanding how usage of CMC tools impacts members’ social interaction, researchers have examined people’s perception of how they use and how they feel about the information and functions conveyed by the CMC tools (Davis, 1989). The Technology Acceptance Model (TAM) is a frequently used framework for exploring people’s technology usage behaviors (Davis, 1989). Since people’s intention of using technologies influences their attitude about using the tools and their perception of how useful the tools are, two primary constructs, perceived usefulness and perceived ease of use, have been identified as determinants of people’s technology acceptance (Mathieson, 1991; Davis, 1989). By adapting TAM to examine students’ online learning experience, Lin (2005) showed how students’ intention of using technologies impacted students’ appropriation behavior. Also, previous study, which explored the relationships among perceived ease of use (PEU), perceived usefulness (PU), SA, and SOC in online courses, found that PEU directly influences SA and PU positively impacts SOC (Tsai et al., 2008a). However, it is not clear if these relationships found in online courses exist in the same way in a teacher CoP.

**Relationships among social constructs of CoP**

Since SOC and SA have been recognized as influential factors determining levels of participation and interaction in online learning environments, studies of online courses have examined how these particular social constructs influence students’ learning satisfaction, course grade, and participation. Because the social constructs are interdependent and intertwined, it makes their contribution difficult to understand without considering the relationships among them. There are rare studies examining the relationships among social constructs as a whole. In a study which established a unified model to explain the relationship among social constructs in online courses, Lin et al. (2006) found that SA was a significant predictor for students’ online learning satisfaction and students’ perceptions of using tools in an online context influence not only their learning satisfaction but also how they used the tools to better achieve learning goals. In another study examining both SA and SOC simultaneously (Tsai, Yang, & Laffey, 2008a), researchers found SOC could be a mediator for the relationship between SA and learning satisfaction, and SA was not a direct predictor for learning satisfaction. When Lin et al. (2006) examined the relationship between SA and learning satisfaction, social connectedness was included as a factor of SA. It is not clear how the relationship between SA and learning satisfaction are influenced by the social connectedness which has a similar meaning to SOC. These two studies examined the relationships between SA and other social constructs, but they did not examine the sub-relationships among SA’s sub-constructs and other social constructs. While there is reason to believe that social constructs influence online learning, more research examining all the factors simultaneously is needed to understand the relationships among these key social constructs and online learning outcomes.
Previous studies have developed reliable instruments to measure SOC and SA in online course contexts but have not been tested in an online CoP. To better understand the social nature of a CoP, there is a need to explore how SOC and SA influence teachers’ social interaction and effectiveness of participating in the online teacher CoP. Thus, this study examined not only the relationships among key social constructs but also relationships among members’ perception of technology acceptance, SA, and SOC.

**Research questions**

The purpose of this study was to understand the social nature of an online learning community of practice. To understand the social nature, the relationships among social constructs were examined via path analysis. The proposed path model (Figure 1) of how well the social constructs of the online learning community explain the effectiveness of professional development in a community was adapted from Tsai et al. (2008a). Below are the two research questions for this study.

(1) How well does the proposed path model explain the relationships among the social constructs of online learning (e.g., SOC, SA, PEU, and PU) and explain community outcomes/effects (satisfaction with NETwork experience (S) and effectiveness of NETwork for Teaching (ET))?

(2) How do members’ perceptions of the sub constructs of social ability influence other social constructs, and how are they impacted by other social constructs?

**Research method**

**Research context and participants**

An online teacher community, NETwork, has been in place using the Sakai course management system to support K-8 science teaching since August 2006. The purpose of NETwork at the University of Missouri is to overcome the current disconnection between pre-service teacher education and in-service teaching practice and to provide pre-service and in-service teachers a collaborative learning environment. The community was established via recruiting members from the field experience courses in the teacher education program. The key features of NETwork include (a) communication tools: synchronous and asynchronous tools, files sharing space, and some notification and social awareness tools; and (b) learning tasks: asynchronous topic discussions, synchronous chat discussion sessions, and teaching resources or lesson plans sharing.

---

1 The represents direct relationship, “E” represents residual error.

* Figure 1. Proposed Path Model
In addition to professors of the teacher education, members of Network include pre-service teacher/old members (POs) who took the field experience courses in Fall 2006 semester, pre-service teacher/new members (PNs) who was taking the courses in Spring 2007 semester, in-service teacher/old members (IOs) who joined the community since Fall 2006, and in-service teacher/new member (INs) who joined the community since Spring 2007. The professors served as a facilitator to guide the discussion and provide relevant resources of teaching. Based on the knowledge and skills learned, the pre-service teachers served as a learner role to discuss how to teach science as well as gain practical experience from what in-service teachers shared about their school teaching. However, NETwork is not just a CoP within an online course. While it did integrate some activities (e.g., topic discussion of teaching methods and resource sharing) for PNs while undertaking their field experience courses, the primary goal of Network is to establish connections that go beyond the pre-service classroom in time and place. Most of the NETwork activities (e.g., topic discussion focusing on K-12 science teaching, synchronous chat discussion session with in-service teachers, and lesson plan and resource sharing) were not necessarily related to what PNs were doing in their courses.

There were a total of 92 student members in NETwork. Table 1 presents the demographic information of 66 members who participated in the final survey.

Table 1. Demographic Information for 66 Participants

<table>
<thead>
<tr>
<th>Demographic Information</th>
<th>Number of Participants</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6</td>
<td>9.1</td>
</tr>
<tr>
<td>Female</td>
<td>60</td>
<td>90.0</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 20</td>
<td>7</td>
<td>10.6</td>
</tr>
<tr>
<td>21-25</td>
<td>55</td>
<td>83.3</td>
</tr>
<tr>
<td>26-30</td>
<td>3</td>
<td>4.5</td>
</tr>
<tr>
<td>&gt;30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>missing data</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Membership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old member (since Fall06)</td>
<td>36</td>
<td>54.5</td>
</tr>
<tr>
<td>New member (since Spring07)</td>
<td>30</td>
<td>45.5</td>
</tr>
<tr>
<td><strong>Teaching Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service teacher (Old member)</td>
<td>30</td>
<td>45.5</td>
</tr>
<tr>
<td>Pre-service teacher (New member)</td>
<td>30</td>
<td>45.5</td>
</tr>
<tr>
<td>In-service teacher (Old member)</td>
<td>6</td>
<td>9.1</td>
</tr>
<tr>
<td>In-service teacher (New member)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Participation of NETwork Discussion</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>participator</td>
<td>42</td>
<td>63.6</td>
</tr>
<tr>
<td>non-participator</td>
<td>24</td>
<td>36.0</td>
</tr>
<tr>
<td><strong>Previous Online Learning Experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Sakai</td>
<td>46</td>
<td>69.7</td>
</tr>
<tr>
<td>In Blackboard</td>
<td>48</td>
<td>72.7</td>
</tr>
<tr>
<td>In other systems</td>
<td>2</td>
<td>3.0</td>
</tr>
<tr>
<td>Missing data</td>
<td>14</td>
<td>21.2</td>
</tr>
<tr>
<td><strong>Messages Posted in Discussion Board (weekly)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than 2 postings</td>
<td>37</td>
<td>56.1</td>
</tr>
<tr>
<td>3-5 postings</td>
<td>24</td>
<td>36.4</td>
</tr>
<tr>
<td>6-8 postings</td>
<td>4</td>
<td>6.1</td>
</tr>
<tr>
<td>8-10 postings</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Hours Login(weekly)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than 1 hour</td>
<td>38</td>
<td>57.6</td>
</tr>
<tr>
<td>1-5 hours</td>
<td>28</td>
<td>42.4</td>
</tr>
<tr>
<td>6-10 hours</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>&gt;10 hours</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Instruments**

Below are the constructs included in the survey. The items for assessing the social constructs were adapted from previous studies. For the social constructs items, participants were asked to rate their agreement with the items on a 7-point Likert-type Scale where 1 represented strongly disagree and 7 meant strongly agree.
**Technology Acceptance.** The 10 items of these constructs were adapted from an online learning experience survey (Yang et al., 2006) based upon Davis’s (1989) technology acceptance instrument. The Cronbach $\alpha$ reliability of PEU and PU assessed in the previous study were .96 and .98 (Tsai, Kim, Liu, Goggins, Kumalasari, & Laffey, 2008).

**SOC.** The 20 items for SOC were adapted from Rovai’s Classroom Community Scale (Rovai, 2002) to measure SOC in an online learning environment. The Cronbach $\alpha$ reliability of connectedness subscale and learning subscale are reported as .92 and .87 (Rovai, 2002).

**SA.** The 18 items of social ability were adapted from a 30-item instrument of online learning experience (Yang et al., 2006). Advances to the original 20-item instrumentation, Yang et al. (2006) found students perceived instructor’s social presence and peers’ social presence differently. Based upon this result, this study examined social presence with instructors (instructor role: professors and mentors, 6 items) and social presence with peers (peer role: pre- and in-service teachers, 6 items) separately. The Cronbach $\alpha$ reliability of SN, social presence with instructor (SPI), and social presence with peers (SPp) were .88, .93, and .91 (Yang et al., 2006).

**S.** 9 items to measure members’ satisfaction with NETwork experience were modified to meet the context of this study from a previous online learning experience study (Laffey et al., 2006; Yang et al., 2006). The Cronbach $\alpha$ reliability of overall satisfaction of online learning were .92 and .87 for learning satisfaction and .89 for course evaluation (Tsai et al., 2008).

**ET.** The 10 items about how members perceive the value of participating in NETwork were developed by the author to address how members feel about their teaching knowledge, skills, and confidence after participating in NETwork. Two expert reviews were conducted for developing these items. The expert with expertise in science education reviewed the items to ensure the questions could address members’ feeling of their improvement/change of the content and pedagogical knowledge in science teaching. After that, the expert with expertise in social computing reviewed the questions to ensure questions were realistic and meaningful to measure members’ changes after participating in an online learning community.

**Data collection and analysis**

To address the research questions a final survey collected at the end of the semester was employed. An invitation with a link to a consent form and online survey was sent to the 92 members, and 66 members (72%) completed the survey. Additionally, path analysis which visualizes the intertwined relationships was employed to analyze the final survey data to quantify the relationships among the social constructs of online experiences in NETwork. To further understand how sub constructs of social ability are related to other social constructs, two further path analyses were conducted to assess how members’ technology acceptance influences the sub constructs of SA and how the sub constructs of social ability impact members’ SOC and satisfaction of participating in NETwork.

**Results**

**Preliminary analysis**

Prior to the path analysis, the assumptions of normality, linearity, multicollinearity, and homoscedasticity were examined and found to be satisfactory. There were no univariate or multivariate outliers found and no cases were excluded. Descriptive statistics and Cronbach’s alpha reliability estimates for the social constructs are presented in Table 2. All social constructs showed satisfactory reliability with Cronbach’s alpha values greater than .80 (Nunnally, 1978).

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>Reliability (# of items)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of Community (SOC)</td>
<td>4.90</td>
<td>.96</td>
<td>.92(20)</td>
</tr>
<tr>
<td>Social Ability (SA)</td>
<td>4.93</td>
<td>1.19</td>
<td>.96(18)</td>
</tr>
<tr>
<td>Social Presence with Peer (SPp)</td>
<td>5.07</td>
<td>1.32</td>
<td>.92(6)</td>
</tr>
<tr>
<td>Social Presence with Instructor (SPI)</td>
<td>4.90</td>
<td>1.34</td>
<td>.91(6)</td>
</tr>
<tr>
<td>Social Navigation (SN)</td>
<td>4.83</td>
<td>1.31</td>
<td>.91(6)</td>
</tr>
</tbody>
</table>
Table 3 presents a correlation matrix of all the social constructs. The results present how social constructs are correlated and provide a basis for making decisions of including or dropping constructs for testing the path model.

### Table 3. Correlations among Social Constructs (N=66)

<table>
<thead>
<tr>
<th>Variables</th>
<th>SOC</th>
<th>SA</th>
<th>SPp</th>
<th>SPi</th>
<th>SN</th>
<th>PEU</th>
<th>PU</th>
<th>S</th>
<th>SE</th>
<th>LS</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of community (SOC)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social ability (SA)</td>
<td>.814**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Presence with Peer (SPp)</td>
<td>.733**</td>
<td>.920**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Presence with Instructor (SPI)</td>
<td>.743**</td>
<td>.910**</td>
<td>.783**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Navigation (SN)</td>
<td>.730**</td>
<td>.879**</td>
<td>.707**</td>
<td>.679**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived ease of use (PEU)</td>
<td>.715**</td>
<td>.720**</td>
<td>.688**</td>
<td>.721**</td>
<td>.539**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness (PU)</td>
<td>.725**</td>
<td>.741**</td>
<td>.653**</td>
<td>.726**</td>
<td>.627**</td>
<td>.865**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction (S)</td>
<td>.771**</td>
<td>.784**</td>
<td>.668**</td>
<td>.733**</td>
<td>.722**</td>
<td>.742**</td>
<td>.810**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site Evaluation (SE)</td>
<td>.702**</td>
<td>.716**</td>
<td>.630**</td>
<td>.700**</td>
<td>.608**</td>
<td>.720**</td>
<td>.761**</td>
<td>.959**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Satisfaction (LS)</td>
<td>.782**</td>
<td>.792**</td>
<td>.660**</td>
<td>.717**</td>
<td>.769**</td>
<td>.718**</td>
<td>.802**</td>
<td>.975**</td>
<td>.872**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness of NETwork for Teaching (ET)</td>
<td>.683**</td>
<td>.722**</td>
<td>.571**</td>
<td>.691**</td>
<td>.694**</td>
<td>.635**</td>
<td>.759**</td>
<td>.900**</td>
<td>.842**</td>
<td>.894**</td>
<td></td>
</tr>
</tbody>
</table>

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

### Examination of the proposed path model

In the final path model, the correlation coefficients of the direct paths range from .23 to .90 and are statistically significant at p < .05 or .001. The R’s means indicate that approximately 80% of the variance of ET is explained by members’ S in the NETwork community. Members’ perceptions of SOC, SA, and the PU explain approximately 72% of the variance of members’ S. Also, SA and PU account for 68% of the variance in members’ SOC. Lastly, members’ perceptions of PEU directly accounted for about 52% of variance of members’ SA and approximately 75% of variance of PU.

Note. 7 point Likert-type Scale where 1 represented strongly disagree and 7 meant strongly agree were used.
Table 4. Model Fit Indices

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>P</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
<th>RMSEA</th>
<th>RMSEA 90% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>N/A</td>
<td>&gt;.05</td>
<td>&gt;.95</td>
<td>&gt;.95</td>
<td>&lt; .10</td>
<td>&lt; .06</td>
<td>90% C. I.</td>
</tr>
<tr>
<td>Results of the Final Model</td>
<td>15.76</td>
<td>.03</td>
<td>.98</td>
<td>.96</td>
<td>.05</td>
<td>.14</td>
<td>.04 ~ .23</td>
</tr>
</tbody>
</table>

Note: N=66 (Hu & Bentler, 1999)

Figure 2. Final Path Analysis Model with R² Values

Figure 3. Mediator Identification

1 * represents $z > 1.96$, $p < .05$; ** represents $z > 3.29$, $p < .001$ statistically significant; —— represents non-significant path, —— represents significant path without decreasing strength; —— represents weaken path with significant value.
Members’ perceptions of SOC did not have a direct impact on their perception of ET as had been expected and included in the initial path model. Also, members’ perception of S was the only influential factor to how they perceived the ET. In the proposed path model, SA had been expected to influence S via the mediation of SOC. However, in the final model, S was directly explained by SA.

Additionally, the mediating relationships among the variables in the final path model were examined. According to Frazier, Tix, and Barron (2004), a mediator is defined as a variable that accounts for the relation between a predictor and a dependent variable. Five potential mediating relationships, including PU→SOC→S, PEU→PU→S, SA→S→ET, and SOC→S→ET, were identified. Based on the examining steps presented in Frazier, Tix, and Barron (2004), a full mediator is identified when the relationship between a predictor and a dependent variable becomes non-significant after adding direct relationship between the mediating variable and the dependent variable. After the examination, PU was found to be a full mediator for the relationship between PEU and S, as well as S was a full mediator for both the relationships between SOC and ET and SA and ET. Additionally, members’ perception of SOC was found to be a partial mediator for the relationship between PU and S. The procedures for establishing the full mediating relationships are illustrated in Figure 3.

**Examination of SA sub-constructs**

To further examine the relationships based on relationships (i.e. SA→SOC, SA→S, PEU→SA) identified in the final path model, SA was replaced with its sub-constructs, including SN, SPI, and SPP. Due to the small sample size, the examinations were conducted separately to meet the required minimum sample size (at least 5 cases per parameter). The relationships among SA’s sub-constructs, PEU, and PU were examined as the first set. Later, a second analysis for SA’s sub-constructs, SOC, and S was undertaken.

**First Set Examination.** In the path analysis for the first examination, two non-significant paths (PEU→SN & PU→SPP) were dropped. A final path model presented in Figure 4 was achieved. According to Hu and Bentler (1999), the indices reported in Table 5 suggest a good model fit.

<table>
<thead>
<tr>
<th>Model Criteria</th>
<th>χ²</th>
<th>P</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
<th>RMSEA</th>
<th>RMSEA 90% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results of the Final Model</td>
<td>2.05</td>
<td>.36</td>
<td>1.00</td>
<td>1.00</td>
<td>.04</td>
<td>.02</td>
<td>.00 – .25</td>
</tr>
</tbody>
</table>

*Note. N=66 (Hu & Bentler, 1999)*

![Figure 4. Final Path Analysis Model (Set 1) with R² Values](image-url)

Table 5. Model Fit Indices (Set 1)

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²</th>
<th>P</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
<th>RMSEA</th>
<th>RMSEA 90% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>N/A</td>
<td>&gt;.05</td>
<td>&gt;.95</td>
<td>&gt;.95</td>
<td>&lt; .10</td>
<td>&lt; .06</td>
<td></td>
</tr>
<tr>
<td>Results of the Final Model</td>
<td>2.05</td>
<td>.36</td>
<td>1.00</td>
<td>1.00</td>
<td>.04</td>
<td>.02</td>
<td>.00 – .25</td>
</tr>
</tbody>
</table>

*represents z > 1.96, p < .05; ** z > 3.29, p < .001 statistically significant;  represent significant direct path,  represents significant correlation paths,  represents variance explained.
In the final path model of set 1, the correlation coefficients of the direct paths range from .24 to .87 and are statistically significant at p<.05 or .001. The R’s means showed approximately 87% of the variance of PU was explained by PEU. Also, PEU and PU explain approximately 54% of the variance of SPI. Members’ perception of PU accounted for 36% of the variance of SN, and PEU explained 46% of the variance of SPP.

Additionally, two possible full mediating relationships among the variables in the set 1 final path model were examined (Figure 5). PU was identified as a full mediator for the relationship between perceived PEU and SN. When adding the direct path between PU and SN, the value of the direct path from PEU to SN not only dropped but also became non-significant. However, PU was found to only partially mediate the relationships between PEU and SPI. After adding the direct path between PU and SPI, the path value between PEU and SPI dropped but remained significant.

Second Set Examination. For the second set of constructs, the path analysis results were used to discard two non-significant paths (SPP→S & SPP→SOC). Although the standardized root mean square error of approximation (RMSEA) was .11 and did not meet the criteria of good fit, the confidence interval of RMSEA was found between .00 and .29 including .05. These findings suggest a marginal fit of the model. The other indices for the goodness of the model fit presented in Table 7 indicate a good model fit for this over-identified model. Overall the data fits the model well. To visualize the relationships among the SA sub-constructs, the final path model is presented the Figure 6.

In the final path model, the correlation coefficients of the direct paths range from .26 to .78 and are statistically significant at p < .05 or .001. The R’s means showed SN, SOC, and SPI explain approximately 68% of the variance of S. Also, SN and SPI explain approximately 65% of the variance of SOC. The correlation between the SA sub-constructs ranged from .68 to .78, which indicated a high correlation among the SA sub-constructs.
Additionally, two possible full mediating relationships among the variables in the set 2 final path model were examined (Figure 7). SOC was found to only partially mediate the relationships between SN and SPi to S. When adding the direct path between SOC and S, the value of the direct path from SN to S remained significant, as well as the path value between SPi and S. Thus, SOC is failed to serve as a full mediator in this path model.

**Figure 7. Mediator Identification (Set 2)**

1 * represents z < 1.96, p < .05; ** represents z < 3.29, p < .001 statistically significant; —— represents significant direct path, ———— represents significant correlation paths, —— represents variance explained.

**Figure 6. Final Path Analysis Model (Set 2) with R² Values**
Discussion

Differing from prior studies (Tsai et al., 2008; Lin et al., 2006) that utilized statistical models to visualize the relationships among social constructs of online learning in courses, an additional variable, ET, was extension in this study. Members’ S was found to be the only direct factor impacting ET, which means members felt the effectiveness of NETwork for their teaching only when they felt satisfied with learning in NETwork. Additionally, S was a mediator for the relationships of SA to ET and SOC to ET, which means members’ SA and SOC influenced ET indirectly. These mediating relationships are new discoveries in the final path model and extend previous models.

Additionally, SA, SOC, and PU directly impact S, meaning members with higher SA and SOC felt much more satisfied with their learning experience, as well as members who felt the Sakai tools were useful for their learning and interaction were much more satisfied with their learning experience. SA and PU account for significant variation in SOC, indicating members had higher SOC when they felt the usefulness of Sakai tools to facilitate their learning and to socially interact with others in NETwork. This result supports the insight that social interaction can be supported by CMC tools if the tools are utilized effectively and members feel the usefulness of the tools for their learning (Tu & Corry, 2003; Lavooy & Newlin, 2003; Tu & McIsaac, 2002).

Further, PEU directly influences SA and PU, meaning that members needed to feel the ease of use of Sakai tools before they can perceive the usefulness of Sakai tools and utilize them to socially interact with others. PU was a mediator for the relationships between PEU and S, indicating PEU contributes to S when they also feel the usefulness of the Sakai tool. This finding is consistent with Hillman, Willis, and Gunawardena (1994) that argued users’ learning is obstructed if they cannot interact easily through the medium/tools. Thus, ease of use of the tools is one of the most basic requirements for establishing an online learning community.

According to Tsai et al. (2008) and Line et al. (2006) that reported SOC fully mediates the relationships between SA and satisfaction in online courses, it was expected that SA would not directly impact S when considering SOC and other social constructs simultaneously. However, this study found a direct relationship between SA and S exists even when adding sense of community in the model and failed to confirm SOC’s fully mediating role for the relationship between SA and S. Additionally, SOC was only a partial mediator for the relationship between PU and S in the present study, which is inconsistent with the findings of Tsai et al. (2008). These inconsistencies might result from the different contexts in this study (an online learning community) and that of Tsai et al. (2008) (online courses).

Going beyond prior studies, this study provided further examination of the interdependent relationships among SA sub-constructs (SN, SPI, and SPP) and other primary social constructs (e.g., SOC, SA, PEU, and PU). The results indicate that PU was significantly associated with SN and SPI. Prior studies (Tsai et al., 2008; Lin et al., 2006) and this study showed no significant relationship between PU and the main construct, SA; however, in the examination of SA sub-constructs, significant relationships of PU to SN and PU to SPI were found. This indicates members who perceived the usefulness of Sakai tools had a better sense of instructors’ social presence and appreciated being able to use what others did as a guide for their own actions. Additionally, PU was found to fully mediate the relationship between PEU and SN. It is possible that a direct influence from PU to SA was not identified in the final path model in this study and that of Tsai et al. (2008). Thus, the direct relationship from PU to SA is suggested for a further examination.

Further, SN and SPI were found to directly influence SOC and S, while SPP did not have any direct impact on SOC and S. Similar to the final path model, SOC was not a full mediator for the relationship between SA and S. SOC was only a partial mediator for the relationships between SA sub-constructs and S (i.e., SN→S and SPI→S). After adding SOC in the mediator identification models, these two relationships (SN→S and SPI→S) weakened but remained significant. In addition to the study of Tsai et al. (2008), which examined the relationships among SOC, S, and SA as a whole, this study provides new insights for understanding how different SA sub-constructs influence SOC and S within a CoP experience.

References


Rovai, A. P. (2002a). Building sense of community at a distance. *International Review of Research in Open and Distance Learning, 3*(1), 1-16.


Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. The Elementary School Journal, 104(2), 127-146.


Cyber Asynchronous versus Blended Cyber Approach in Distance English Learning

Zi-Gang Ge
School of Network Education, Beijing University of Posts and Telecommunications, China // shouzhou11@126.com

(Submitted November 6, 2010; Revised February 12, 2011; Accepted March 26, 2011)

ABSTRACT
This study aims to compare the single cyber asynchronous learning approach with the blended cyber learning approach in distance English education. Two classes of 70 students participated in this study, which lasted one semester of about four months, with one class using the blended approach for their English study and the other only using the cyber asynchronous approach. Students’ final scores were collected and processed at the end of the semester. The null hypothesis was that there would be no significant differences in the outcomes of the two approaches. But the data obtained repudiates the null hypothesis and shows that although both approaches improved students’ performance, the blended approach could bring a significantly better result for adult e-learners in their English study than the single cyber asynchronous approach. The questionnaire survey at the end of the study indicates that cyber synchronous learning can provide students with some unique help which cannot be obtained in cyber asynchronous learning.

Keywords
Cyber synchronous learning, Cyber asynchronous learning, Blended cyber approach, Online courses

Introduction

With the fast development of broadband Internet and computer technologies, online courses and thus cyber asynchronous learning have been employed more and more often for exchanging information between instructors and students, and between students and their peers (Hew et al., 2010). Cyber asynchronous learning allows students to have more freedom to conduct their learning process without the constraints of time and space. Meanwhile, the more traditional cyber synchronous learning through TV or satellite broadcasting or some other teaching systems is now gradually disappearing. At least, this is true in China, especially with Chinese adult e-learners (Ge, 2011). Cyber synchronous learning in China usually requires that students conduct their learning by attending real-time lectures through some sort of videoconference system. After all, most adult e-learners have regular jobs and may not meet the time requirements of cyber synchronous learning. Besides, more and more e-learning institutions have begun to encourage their students to utilize the cyber asynchronous learning environment instead of waiting for cyber synchronous instruction.

Cyber asynchronous learning supporters cry out the advantages of this learning approach. They point out that cyber asynchronous learning allows students to study through emails, blogs, etc., and students can make out their own schedule, without live interaction with the instructor (Kruse, 2004). Cyber asynchronous learning allows learners to determine whether or when to participate in course activities without considering whether other learners or the course instructor is present in the virtual learning system. So the obvious advantage of cyber asynchronous learning is convenience. As most adult e-learners are often busy with their jobs, this kind of convenience is really very precious for them. Besides, students often have to rely on themselves in a cyber asynchronous learning environment, for their teachers cannot always wait for them online. This actually can improve one’s personal ability in learning (Robert & Dennis, 2005). Asynchronicity can also enhance educational assessment of students’ learning processes. Cyber asynchronous teaching platforms normally can keep records of a student’s online learning activities, including discussion threads and his or her interactions with peers and/or the instructor, which can be an important source of data for the assessment of the learner (Tanimoto et al., 2002; Shi et al., 2006; Hew et al., 2010). Moreover, these records can enable the student to review his or her learning activities at any time, and this kind of reviewing and reflection can help enhance the student's higher level learning, such as analysis and evaluation (Newman et al., 1997). On the other hand, cyber asynchronous learning does not necessarily mean that real-time interaction cannot happen. A good case in point is a threaded discussion, which is asynchronous in nature but also involves intensive interaction. Seeing these advantages, many e-learning institutions have begun to develop online learning courses, which are the most important part of a cyber asynchronous learning environment. In addition, online learning systems may also enable them to employ fewer faculties and thus cut their costs.
Meanwhile, some other scholars hold different opinions. They have shown great concern about the learning outcomes of cyber asynchronous learning. They are wondering whether learners can really follow the teaching plans set by their instructors. They think that the quality of instruction and the ability of students to master courses should be observed and assessed as new technologies are involved (DiPiro, 2003). The most troublesome problem in a cyber asynchronous learning environment is that students may have few chances or little desire to interact with their peers or teachers. Some scholars claim that limited student contribution is a persistent and widespread problem in cyber asynchronous discussion (Hewitt, 2005; Hew et al., 2010). Without enough contribution or interaction, students’ learning can hardly produce satisfactory results and students will often feel isolated and out of the learning communities (Haythornthwaite & Kazmer, 2002).

It is generally believed that synchronous interaction is essential to second language acquisition (SLA) (Lee, 2002). The cyber synchronous learning environment can duplicate the capabilities found in a physical face-to-face classroom (Keegan et al., 2005; Shi et al., 2006). As summarized by Desmond Keegan and some other scholars (2005), the benefits of cyber synchronous learning include “(i) the familiarity of the classroom model, (ii) learners receive immediate feedback from other learners and the leader, (iii) the ability to create content quickly in the classroom”. The study of Pfister (2005) indicates that synchronous net-based discourses among learners, or among learners and instructors can greatly improve understanding of complex subject matters. Pilkington and Walker’s work (2003) even suggests that non-native English speakers working collaboratively through a virtual learning environment can outperform face-to-face students in group work on the same course.

But scholars on the opposite side have pointed out the disadvantages of cyber synchronous learning. They say the instructor and the students may often feel pushed by the time limit in a cyber synchronous environment. The instructor wants to cover all he or she has prepared and a student will try hard to grasp all the instructor has covered. So the focus of cyber synchronous learning is often on quantity rather than quality (Hrastinski, 2008). Lee’s study (2002) also indicates that cyber synchronous exchanges among non-native speakers of English tend to encourage fluency rather than accuracy. Lee insists that synchronous online exchanges should maintain a balance between function, content, and accuracy (ibid.). Besides, cyber synchronous learning requires all students to be available at a certain time for the synchronous videoconferencing, and this is often regarded as the major disadvantage of this learning approach (Mirza, 2007).

The present study

Everything has two sides. Now, we have three choices. The first is to adhere to the more traditional cyber synchronous learning approach and shrug off the more popular cyber asynchronous learning approach, the second is just the opposite and the third is to keep both and combine the two together. As to the first choice, normally we will cast it away, as it is obviously against the trend of modern educational technologies. As a result, we have two choices left. Evidence suggests that learners often prefer the blended approach that includes both forms (Gregory, 2003). The blended approach can fulfill different types of needs and foster the participation of people with different capabilities and competencies (Ligorio, 2001). The current e-education practice in China, however, tends to reduce the use of cyber synchronous means and let the cyber synchronous learning approach be dominant (Ge, 2011). So the question is whether there is a possibility that one (the single cyber asynchronous approach) is better than two (the blended cyber approach)? This paper aims to address this question through an empirical study.

The researcher of this paper expected all the participants would improve their English abilities after this 4-month experiment. Participants would take a post-experiment test to examine their learning outcomes and they would also be required to respond to a questionnaire survey at the end of the study, with the aim to obtain their perceptions of the two learning approaches. The data analysis and the survey would be conducted to address the following questions:

- Did all the participants improve their English abilities and which approach could bring a better result?
- What attitudes did the participants have toward the two approaches?
- What problems might occur in the participants’ study through the two approaches?

Methods

Participants

The participants consisted of two classes of adult e-learners (23-35 years old) who came from a network education program.
college of a university situated in Beijing. The learners were newcomers in the college and they did not have any previous experiences in e-learning. The course instructor, however, is an experienced professional in e-education, who had been teaching English courses for about 10 years. Consent was obtained from the college and the participants prior to the conduct of this study, which lasted one semester of about 4 months. Class 1 was composed of 34 students (24 males and 10 females) and Class 2, 36 students (22 males and 14 females). All the students were full-time job holders and about 10 percent in each class were high school graduates, with the rest below this level. Both classes were majoring in Computer Science. Class 1 was taken as the control group for the study and Class 2, the experimental group.

Procedure

Both classes were introduced to the asynchronous online course of College English Level 2. The online course provided students with prerecorded video lectures, downloadable materials, online quizzes and exercises, an online discussion forum, etc.

The students could access the online course by using their office or home computers according to their own time arrangement. The course could also be accessed through a Wap browser in mobile phones, which made mobile learning a new choice for the students. The students could study on their own by watching the prerecorded video lectures and doing the online exercises. They might also want to communicate to others by reading or writing posts on the online forum. The teacher would regularly organize some asynchronous online discussion activities on the forum.

Only one English teacher was there to cater for the English teaching and assistance to the two classes, whether synchronously or asynchronously. The online materials were enough for students to complete the course and pass the final examination. There was no special requirement as to how to use these materials, so students could make their own judgments about when or how to use them. The cyber asynchronous learning system could track the online activities of the students such as the frequency of their participation. Students were required to study online for at least four hours a week, and this requirement was to ensure that they really carried out their study. After all, online activities are easier to inspect than offline ones.

Students in Class 2 would also attend some synchronous videoconference classes. The synchronous videoconference system is called Webex (http://www.webex.com.cn), which is a very powerful videoconference system that can transmit audio and visual signals across the Internet. The cyber synchronous classes were organized into eight 3-hour lectures. There would be two 15-minute intervals during each lecture, that is to say, each lecture would be divided into three sessions with each session lasting about 50 minutes. There would be two lectures in each month. In cyber synchronous classes, the teacher would spend about half of the time summarizing the most important knowledge points of the course such as grammatical structures and the rest of the time would be used for interaction with the students. The interaction would normally include some questions for the students to answer. Students were encouraged to volunteer to answer the questions. If there were no volunteers, the teacher would randomly choose a student to talk to, but would try to ensure every student to have a chance to participate in the activities.

Students might respond by typing on keyboards or talking directly with the teacher through the microphones connected to their computers. The interaction between the instructor and the students normally included three or more turns. A typical interaction followed the pattern shown by Table 1.

| Turn 1: | Instructor (speaking through the microphone): Greeting (for example, “Hi, Wang Lei (student’s name). Glad to meet you online. I have some questions for you.”) |
| Turn 2: | Student (speaking or typing): Greeting (for example, “Hi, teacher! Happy to talk with you online.”) |
| Turn 3: | Instructor: Asking a question |
| Turn 4: | Student: Answering the question |
| Turn 5: | Instructor: Evaluating the student’s answer |
Student: Responding to the evaluation
The interaction may end here, or the instructor may continue to ask the student another question, and then the interaction goes on.

The interaction could involve various kinds of topics (mostly related to the textbook). Students could speak Chinese or English during interaction, as most of them were poor in spoken English. These activities were to make sure that the students had necessary interaction with their teacher. The teacher had a microphone and a video camera connected to his computer. All the students were required to equip their computers with a microphone, with a video camera as an option. This technological requirement was to ensure the effectiveness of the cyber synchronous interaction. But of course, if the students didn’t want to talk directly to the teacher, then they could use keyboards to type out their messages.

There would be ten obligatory online assignments for the two classes to do. Nine were about multiple choice questions and one was essay-writing. The multiple choice questions were checked by a built-in function of the online course and the essay-writing assignment was marked by the teacher. At the end of the semester, all the students would attend a final examination.

The students’ scores of the final examination (final scores) would be processed and analyzed to see the outcomes of the two learning approaches. Besides, two different questionnaires were distributed to the two classes respectively after the final examination. The purpose was to obtain their perceptions of the two approaches.

At the beginning of the study, all the students’ English scores of the entrance examination (entry scores) were processed for Levene's Test for Equality of Variances in SPSS. See Table 2.

<table>
<thead>
<tr>
<th>class</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Levene's Test for Equality of Variances</th>
<th>F</th>
<th>Sig.</th>
<th>Sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>class1</td>
<td>34</td>
<td>49.7941</td>
<td>11.00628</td>
<td></td>
<td>2.543</td>
<td>.115</td>
<td>.327</td>
</tr>
<tr>
<td>class2</td>
<td>36</td>
<td>47.3333</td>
<td>9.83579</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From table 2, we can see the standard deviations of both classes are very large. This shows that the students within the two classes were very different in their English levels. On the other hand, the mean difference (49.7941 - 47.3333 = 2.4608) between the two classes is very small, which shows that the two classes were similar in the overall English level, with Class 1 being better. Levene's Test for Equality of Variances shows that the scores between the two classes have equal variances. (F = 2.543, p = 0.115 > 0.05). The 2-tailed p-value is 0.327, which means that there was no essential difference between the scores of the two classes and that the two classes were suitable for the study.

Null hypothesis

The null hypothesis of this study is that there are no significant differences between the final scores of those doing their study with the single cyber asynchronous learning approach (Approach 1) and those with the blended cyber approach (Approach 2).

Results

All the students of the two classes completed the course, so their final scores were all valid for the following analysis. Their final scores were processed for an analysis of covariance in SPSS. See Table 3-4.

<table>
<thead>
<tr>
<th>class</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>class1</td>
<td>34</td>
<td>58.2059</td>
<td>9.95377</td>
</tr>
<tr>
<td>class2</td>
<td>36</td>
<td>62.8056</td>
<td>8.46444</td>
</tr>
</tbody>
</table>
Table 3 shows us the following information:

Compared with Table 2, the means of the final scores of both classes have been greatly improved. Class 1 has increased by 8.4118 \((58.2059 - 49.7941 = 8.4118)\), and Class 2 has increased by 15.4725 \((62.8056 - 47.3333 = 15.4725)\). This improvement shows that both Approach 1 and Approach 2 have exerted a positive impact on students’ learning, and Approach 2 has produced a better result.

The mean difference of the entry scores shows Class 1 was better than Class 2 in the overall English level, but the mean difference of the final scores just shows the opposite. This reverse of the relationship indicates that Approach 2 has exerted a more positive impact on students’ learning than Approach 1.

The standard deviations of the final scores of both classes have decreased (Class 1 decreased by 1.05251, and Class 2 by 1.37135). This shows that the difference in English levels within the two classes has been narrowed.

**Table 4. Test of between-subjects effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>4243.239(^a)</td>
<td>2</td>
<td>2121.619</td>
<td>74.662</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>1922.919</td>
<td>1</td>
<td>1922.919</td>
<td>67.669</td>
<td>.000</td>
</tr>
<tr>
<td>Entry score</td>
<td>3873.294</td>
<td>1</td>
<td>3873.294</td>
<td>136.305</td>
<td>.000</td>
</tr>
<tr>
<td>class</td>
<td>702.117</td>
<td>1</td>
<td>702.117</td>
<td>24.708</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>1903.904</td>
<td>67</td>
<td>28.416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>262970.000</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>6147.143</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Dependent Variable: final score. R Squared = .690 (Adjusted R Squared = .681)*

Table 4 shows that both the independent variable, namely the teaching approaches \((F = 24.708, P = 0)\) and the covariate, namely the entry scores \((F = 136.305, P = 0)\) have brought about significant differences in the final scores.

If the effect of the covariate is excluded, then we can get the adjusted effects of the independent variable on the final scores. See Table 5-6.

**Table 5. Estimated marginal means**

<table>
<thead>
<tr>
<th>class</th>
<th>Mean</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>class1</td>
<td>57.289(^a)</td>
<td>.918</td>
<td>55.458</td>
</tr>
<tr>
<td>class2</td>
<td>63.671(^a)</td>
<td>.892</td>
<td>61.892</td>
</tr>
</tbody>
</table>

*Note. Dependent Variable: final score. Covariates appearing in the model are evaluated at the following values: score1 = 48.5286.*

**Table 6. Pairwise comparisons**

<table>
<thead>
<tr>
<th>(I) class</th>
<th>(J) class</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig. (^a)</th>
<th>95% Confidence Interval for Difference (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>class1</td>
<td>class2</td>
<td>-6.382(^*)</td>
<td>1.284</td>
<td>.000</td>
<td>-8.945</td>
</tr>
<tr>
<td>class2</td>
<td>class1</td>
<td>6.382(^*)</td>
<td>1.284</td>
<td>.000</td>
<td>3.819</td>
</tr>
</tbody>
</table>

*Note. Dependent Variable: final score
Based on estimated marginal means
* The mean difference is significant at the .05 level.
\(^a\) Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments)

Table 5 shows that the adjusted means for Class 1 and Class 2 are 57.289 and 63.671, and Table 6 shows that the mean difference is significant \((P = 0)\) at the 0.05 level, which implies that the two approaches can exert significantly different impact on the final scores when the effect of the covariate is excluded. Hence, the null hypothesis is repudiated.
Questionnaires

The questionnaires were distributed and collected through e-mails after the final examination. The students were told that their answers to the questionnaires would not influence their course grades, which was to ensure that the students would express their real attitudes toward their learning experience. Besides, the collected questionnaires were cross-checked by two teachers. All these were to ensure the validity and reliability of the questionnaire data. All the students of the two classes responded to the questionnaires. Appendix A is the calculation of the questionnaire for Class 2, and Appendix B is about the questionnaire for Class 1.

Questions 1-5 in Appendix A are about content delivery of the course and the rest are about course evaluations. Some information can be concluded from Appendix A:

Most of the students (75%) thought they needed the cyber synchronous classes and only a few students (16.7%) thought they could do their learning without the cyber synchronous classes. This shows that adult e-learners still preferred the more traditional learning mode and most of them might lack the confidence in self-studying in the cyber asynchronous environment. Although they could improve a lot by using the cyber asynchronous learning approach (shown by Table 3), most students might still want to have some cyber synchronous communication with their teacher or peers.

About 83.3% students thought each cyber synchronous session should not exceed 30 minutes, 13.9% students thought the time should be confined between 30 to 50 minutes, and only 2.8% students wanted longer sessions. This is possibly because the English level of most of the e-learners was so low that they might not take in too much in a long session.

Most students wanted the teacher to spend more time on vocabulary, grammar and exercises. An examination of their test papers showed that most of the students were weak at vocabulary and grammar, so they might expect the cyber synchronous classes to solve these problems. A relatively small group of students would spend time in reading articles in the text book. Some of them reported in the open question (Question 9) that some articles in the textbook were hard for them to understand, so they wanted the instructor to explain the articles in detail. Still fewer students would like to have listening and speaking practices. This might reveal another weakness in their English study. They were so poor in English listening and speaking abilities that they did not want to have such practices.

Surprisingly, most students (41.7% + 52.7% = 94.4%) would like to have more turns when interacting with the instructor. This seems to contradict their responses to question 3 that they did not want to focus more on listening and speaking. But a careful examination of their responses to Question 9 revealed that they cherished the rare chances to communicate with the instructor, though most of them had no confidence in spoken English. Besides, they were allowed to speak Chinese during the interaction, which might also encourage their participation.

Most of the students (44.4% + 16.7% + 11.1% = 72.2%) thought they were not active in the cyber synchronous classes. This shows the real side of adult e-learners with low English level. Adult e-learners often lack confidence in their English (whether spoken or written), so they are often afraid of showing their English abilities in public. But there were a few students (10 students) who were very active in the cyber synchronous classes. This is probably because these students had more confidence in their English. Although most of the students were not active in the classes, they might have paid attention to the interaction between the teacher and those active learners. So they, in some sense, were auditing the classes. This might also have exerted a positive effect on their study.

As to the organization of the cyber synchronous classes, most students (58.3% + 16.7% = 75%) held a positive attitude. But about 25% students did not approve the organization.

About 83.4% (72.2% + 11.2%) students thought the cyber synchronous classes had improved their English level, but still a few students thought the classes had no positive effect on their English study.

All of the students thought they had obtained some unique knowledge in the synchronous classes. This indicates that the single cyber asynchronous learning approach could not fully meet the needs of e-learners.
Appendix B shows Class 1’s responses to the questionnaire concerning their cyber asynchronous learning experiences.

Most students (76.5%) thought they could not do the study only through asynchronous means. This is echoed by their responses to Question 3 and 8, as most of them expressed their willingness to interact synchronously with the instructor and all of them wanted to attend some cyber synchronous classes.

As to the use of the online discussion forum, most of them (55.9% + 35.3% = 91.2%) contributed too little. We were disappointed that 61.8% students had never written or answered any post in the online discussion forum. But some statements in the open question (Question 9) show that although they contributed nothing to the forum, they had actually read some of the posts, which, in some sense, is also a form of participation.

When asked who or what they would turn to when in trouble with study, most of them (61.8%) chose to ask for the instructor’s help. Only a few would take advantage of the Internet. This finding is contrary to our initial thought that these Computer Science majors would definitely prefer to use sources on the Internet in their study.

As to the use of the online teaching materials, a relatively large number of students (17.6% + 20.6% = 38.2%) did not make full use of the materials, though most of the students had downloaded (82.4% + 8.8% = 91.2%) the materials and thought (44.1% + 23.5% = 67.6%) these materials were enough for their study.

Discussion

The present study aims to compare the single cyber asynchronous learning approach with the blended cyber approach in distance English learning. The popularity of cyber asynchronous learning in China may surprise some scholars in other parts of the world, as in their mind it is cyber synchronous learning that has become more and more popular with the improvements in technology and increasing bandwidth capabilities (Kinshuk & Chen, 2006). The e-learning industry in China is somewhat different from that in some other countries. It is common for China’s e-learning institutions to enroll too many students, which may exceed their teaching capabilities but can bring them more profits in money. Thus the student-faculty ratio is often very large (Ge, 2009), and they have to rely more on asynchronous means in teaching.

The investigation here compares the single cyber asynchronous learning approach with the blended cyber approach. This kind of comparison may shed some light on the study in this field, as previous studies usually compared synchronous learning and asynchronous learning separately.

The result of this study shows that the blended approach can bring a better outcome. This finding indicates that the single cyber asynchronous approach lacks something in the blended cyber approach. Cyber synchronous learning obviously has brought something new to the asynchronous one. In cyber synchronous classes, students will have a chance to interact directly with the teacher and with their peers and the teacher can also assign a lot of activities for the students to do. Many of these activities such as topic debates are not so easy to carry out asynchronously. Moreover, students will develop a sense of belonging in cyber synchronous learning but not feel isolated and out of the learning community, which is typical of an asynchronous e-learning environment (Haythornthwaite & Kazmer, 2002). This conclusion is consistent with previous observations that asynchronous and synchronous e-learning complement each other and “the combination of these two types of e-learning supports several ways for learners and teachers to exchange information, collaborate on work, and get to know each other” (ibid.). The conclusion is also confirmed by the questionnaire survey conducted at the end of the study, as most students thought they had obtained from the cyber synchronous experience some knowledge which was missing from the asynchronous learning environment and those working only with asynchronous means also demanded some cyber synchronous learning classes.

The questionnaire survey for Class 2 shows that most students seemed not so active in the cyber synchronous classes. This finding is partly inconsistent with the predictions of scholars such as Haythornthwaite and Kazmer (2002), Robert and Dennis (2005). They predicted that cyber synchronous communication could increase students’ psychological arousal and motivation. The reason of the inactivity of these students may lie in the fact that they were so poor in English that they had no confidence in taking part in the synchronous activities. On the other hand, only
10 students were very active in the cyber synchronous classes. Their entry scores and final scores indicate that these students were good at English. They were the so-called high-ability students. So they normally had more confidence in their English, and they possibly wanted to show off their abilities among their peers. If so, then this finding can verify the conclusion of Hrastinski (2008) that the focus of synchronous communication is often on quantity rather than on quality—that is, trying to say something quickly because “someone else will say what I was going to say.”

Another finding of the survey is that most students would pay more attention to the teaching of grammar and vocabulary but not listening, reading or writing in the cyber synchronous classes. Some students’ statements in the open question reveal that they often felt very weak at grammar and vocabulary, which are two basic requirements for language learning, so they wanted the instructor to spend more time on these two parts. They tended to believe that their listening, reading or writing abilities were impossible to get great improvement due to the time limit of the cyber synchronous classes. This may confirm previous observations that cyber synchronous learning is more appropriate for less complex information exchanges (Hrastinski, 2008), as grammar and vocabulary are generally easier than listening, reading and writing in language learning, but may contradict some other findings that cyber synchronous learning may greatly enhance understanding of complex subject matters (Pfister, 2005). The differences in these findings may lie in the differences in subject matters and participants.

The survey also shows that most students would prefer short cyber synchronous teaching sessions. This finding may also be useful for future course planning.

As to the questionnaire survey toward Class 1, a prominent finding is that most of the students did not make full use of the cyber asynchronous resources and means, such as the downloadable materials and the online discussion forum. Research of some other scholars also has the same finding (Wan & Johnson, 1994; Guzdial, 1997). This finding may in part explain why Class 1’s performance on the final examination was not as good as that of Class 2. A careful examination of the posts in the online forum shows that the students would respond more to topics related to assignments and final examinations. This finding is similar to prior studies by Fung (2004), and Khan (2005), which indicate that students’ interest or familiarity with the topic being discussed can make them contribute more in an online discussion. But language learning should not always center on assignments or examinations. The instructor needs to direct students’ learning to overall language acquisition.

The findings of this study generally suggest that the blended cyber approach is a better choice for distance English learning, as asynchronous and synchronous e-learning can cater to different needs of e-learners, which are summarized in Table 7.

<table>
<thead>
<tr>
<th>Table 7. Comparison of the benefits of asynchronous and synchronous e-learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asynchronous e-learning</td>
</tr>
<tr>
<td>•Promoting cognitive participation</td>
</tr>
<tr>
<td>•Suitable for discussing complex issues</td>
</tr>
<tr>
<td>•Without time and space constraints</td>
</tr>
<tr>
<td>•Increasing e-learners’ ability to process information</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note. Adapted from Hrastinski’s model (2008)

Conclusions

The findings of the study indicate that the blended cyber approach can bring a better result for adult e-learners in their English study than the single cyber asynchronous approach. Many Chinese adult e-learners are very poor in English, and they may still need cyber synchronous learning to offer them additional help. Many of them cannot obtain satisfactory results by only doing cyber asynchronous learning. So a cyber synchronous learning environment should still be provided to them, although this may cost distance education providers more money.

Cyber asynchronous and cyber synchronous learning can complement each other in teaching and learning. Cyber asynchronous learning can promote e-learners’ cognitive participation, as asynchronous communication can often
increase one’s ability to process information (Robert & Dennis, 2005). Cyber asynchronous learning is more suitable for discussing complex issues, in which time for reflection is needed (Hrastinski, 2008). The obvious advantage of cyber asynchronous learning is that learners can often carry out their study without the constraints of time and space. On the other hand, cyber synchronous learning can promote e-learners’ personal participation, as it is argued that synchronous communication is “more like talking” compared with asynchronous communication and students will become more motivated (ibid.). Besides, students will develop a sense of belonging, as much interaction between instructors and students and between peers is involved in synchronous communication. In addition, cyber synchronous learning is more suitable for discussing less complex issues, because quick feedback is expected in a cyber synchronous class and there is often no much time for thinking and reflection. A more desirable English e-learning course should fully consider the different benefits of the two learning modes, so that it can meet e-learners’ various needs and solve different problems in teaching and learning.

There are some limitations of this study that need to be recognized. First, the experiment lasts only one semester and thus cannot be considered as a longitudinal study. The data collected in one semester may not fully reflect the real situations of the two approaches. More definitive conclusions might have been drawn if the study had been conducted over a longer period of time. Second, it is indicated that students’ perceptions of cyber asynchronous and cyber synchronous learning were collected at the end of the study, so there was no time for adjusting the organization of the two learning modes. What results would have come up if there had been some adjustment? All these questions need to be explored by further research.

References


### Appendix A

**Responses concerning synchronous learning classes**

<table>
<thead>
<tr>
<th>Questions</th>
<th>No</th>
<th>Have no idea</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do you think you can do the study in the asynchronous environment without attending the cyber synchronous classes?</td>
<td>75% (27 students)</td>
<td>8.3% (3 students)</td>
<td>16.7% (6 students)</td>
</tr>
<tr>
<td>2. How long should each cyber synchronous learning session last?</td>
<td>Less than 30 minutes</td>
<td>between 30 minutes to 50 minutes</td>
<td>More than 50 minutes</td>
</tr>
<tr>
<td></td>
<td>83.3% (30 students)</td>
<td>13.9% (5 students)</td>
<td>2.8% (1 student)</td>
</tr>
<tr>
<td>3. What should the teacher spend more time on in the cyber synchronous classes? (more than one answer can be selected)</td>
<td>vocabulary: 83.3% (30 students)</td>
<td>grammar: 83.3% (30 students)</td>
<td>Articles: 41.7% (15 students)</td>
</tr>
<tr>
<td>4. The appropriate number of turns for an interaction should be:</td>
<td>Less than 3 turns</td>
<td>Between 3 to 5 turns</td>
<td>More than 5 turns</td>
</tr>
<tr>
<td></td>
<td>5.6% (2 students)</td>
<td>41.7% (15 students)</td>
<td>52.7% (19 students)</td>
</tr>
<tr>
<td>5. Were you actively involved in the cyber synchronous activities?</td>
<td>Yes: 27.8% (10 students)</td>
<td>Not too much: 44.4% (16 students)</td>
<td>Very little: 16.7% (6 students)</td>
</tr>
<tr>
<td>6. Do you think the course was well organized?</td>
<td>SA: 58.3% (21 students)</td>
<td>A: 16.7% (6 students)</td>
<td>N: 13.9% (5 students)</td>
</tr>
<tr>
<td>7. Do you think the cyber synchronous lectures have improved your English level?</td>
<td>SA: 72.2% (26 students)</td>
<td>A: 11.1% (4 students)</td>
<td>N: 11.1% (4 students)</td>
</tr>
<tr>
<td>8. Do you think you have learned something you can never obtain through the asynchronous approach?</td>
<td>SA: 83.3% (30 students)</td>
<td>A: 16.7% (6 students)</td>
<td>N: 16.7% (6 students)</td>
</tr>
<tr>
<td>9. Write anything you would like to comment on the cyber synchronous learning experiences:</td>
<td>An open question</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SA=strongly agree; A=agree; N=neutral; D=disagree; SD=strongly disagree*
## Appendix B

Responses concerning asynchronous learning experiences

<table>
<thead>
<tr>
<th>Questions</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do you think you can do the study only through asynchronous means?</td>
<td>No 76.5% (26 students) Have no idea 5.9% (2 students) Yes 17.6% (6 students)</td>
</tr>
<tr>
<td>2. How often do you write or answer a post in the online discussion forum?</td>
<td>None in one week 61.8% (21 students) One to three times in one week 29.4% (10 students) More than three times in one week 8.8% (3 students)</td>
</tr>
<tr>
<td>3. Do you want some live online interactions (such as chatting) with the instructor?</td>
<td>SA 82.4% (28 students) A 8.8% (3 students) N 8.8% (3 students) D 8.8% (3 students) SD</td>
</tr>
<tr>
<td>4. Who or what do you turn to when you have trouble with English study?</td>
<td>The instructor 61.8% (21 students) Classmates or friends 26.5% (9 students) Internet or books 11.7% (4 students)</td>
</tr>
<tr>
<td>5. Have you downloaded the teaching materials?</td>
<td>All of them 82.4% (28 students) Most of them 8.8% (3 students) Very little 5.9% (2 students) never 2.9% (1 students)</td>
</tr>
<tr>
<td>6. Have you frequently used the teaching materials in your study?</td>
<td>All of them 14.7% (5 students) Most of them 47.1% (16 students) Very little 17.6% (6 students) never 20.6% (7 students)</td>
</tr>
<tr>
<td>7. Do you think the teaching materials are enough for your study?</td>
<td>SA 44.1% (15 students) A 23.5% (8 students) N 5.9% (2 students) D 17.6% (6 students) SD 8.9% (3 students)</td>
</tr>
<tr>
<td>8. Do you want to attend some cyber synchronous classes?</td>
<td>SA 88.2% (30 students) A 11.8% (4 students) N</td>
</tr>
<tr>
<td>9. Write anything you would like to comment on the asynchronous learning experiences:</td>
<td>(An open question)</td>
</tr>
</tbody>
</table>

*Note. SA=strongly agree; A=agree; N=neutral; D=disagree; SD=strongly disagree*
Adaptively Ubiquitous Learning in Campus Math Path

Shu-Chuan Shih*, Bor-Chen Kuo and Yu-Lung Liu1

Graduate School of Educational Measurement and Statistics, National Taichung University, Taiwan, 140 Min-Shen Road, Taichung 40306, Taiwan // 1Department of Computer Science and Information Engineering, Asia University, 500 Liufeng Rd., Wufeng, Taichung 41354, Taiwan // ssc@mail.ntcu.edu.tw // kbc@mail.ntcu.edu.tw // liu720402@hotmail.com

*Corresponding author

(Submitted November 6, 2010; Revised February 12, 2011; Accepted March 26, 2011)

ABSTRACT

The purposes of this study are to develop and evaluate the instructional model and learning system which integrate ubiquitous learning, computerized adaptive diagnostic testing system and campus math path learning. The researcher first creates a ubiquitous learning environment which is called “adaptive U-learning math path system”. This system enables students to learn math during their daily campus life with mobile devices beyond web-based education with desktop computers. Moreover, this system can support the adaptive testing and real-time computer-based adaptive remedial instruction after students complete the situated learning on campus math path using ubiquitous technology. Next, a quasi-experiment research is conducted to explore the instructional effectiveness of this system. The 118 subjects are selected from fifth-grade classes in Taiwan. Experimental results indicate that the proposed system can enhance mathematical achievement and the effect of remedial instruction. Furthermore, students’ mathematical connection ability can be improved by the proposed instructional model and learning system. Finally, conclusions for mathematics learning are discussed.

Keywords

Ubiquitous learning, Computerized adaptive diagnostic testing system, Math path, Situated learning

Introduction

Recently, many academics are espousing the merits of situated learning. In situated learning of mathematics, a math path is a common method that bridges the gap between formal learning and the places where the learning is to be applied. A math path includes a series of mathematical learning activities that are designed around campuses or communities, using sports fields, trees, or school gates. These activities make math meaningful by providing students with problems and examples demonstrating its applications in environment and everyday life. In other words, students will develop and consolidate key concepts and skills of mathematics by solving authentic, real-world problems on the math path.

There are two major advantages for the math path learning. The first is to help students understand and value mathematics. The math path will provide an opportunity for participants to be active learners, it will provide a context for the learning of mathematics, and will provide a safe, non-threatening environment in which to understand how math is involved in environment and everyday life. The second purpose is to gain awareness of the connection of concepts in mathematics. Participants in the math path will become aware and understand the mathematics all around them embedded in the surrounding environment of the campus.

However, there also are some restrictions in the design of a traditional math path. For example, in paper-and-pencil based problem solving, it is difficult to immediately share and record students’ processes of solving problems in a traditional math path. Furthermore, when students encounter difficulties when problem solving outdoors, the teacher is usually unable to teach each student or support available resources according to individual needs in the right time and right place, not to mention there is a lack of individual assessment and remedial instruction after math path learning.

With the rapid development of wireless communication and sensor technologies, ubiquitous learning (U-learning) or pervasive learning has become a promising solution to educational problems (Chen, Chang, & Wang, 2008; Chen, Kinshuk, Wei, & Yang, 2008; Chiuou, Tseng, Hwang, & Heller, 2010; Chu, Hwang, & Tsai, 2010; Hwang, Chu, Shih, Huang, & Tsai, 2010; Hwang, Kuo, Yin, & Chuang, 2010; Hwang, Tsai, & Yang, 2008; Kuo, Hwang, Chen, & Wang, 2007; Laine, Islas Sedano, Vinni, & Joy, 2009; Liu & Chu, 2010; Liu, Tan, & Chu, 2009; Si, Weng, & Tseng, 2006; Syvänen, Beale, Sharples, Ahonen, & Lonsdale, 2005; Yang, 2006). In previous literatures, some people use...
the words “pervasive” and “ubiquitous” as synonyms, but some papers show that there is a slight difference in two learning environments, such as Lyytinen and Yoo (2002), Ogata and Yano (2004). According to Ogata and Yano (2004), ubiquitous learning has integrated high mobility with pervasive learning environments. In this study, we define U-learning as a learning paradigm which takes place in a ubiquitous computing environment that enables anyone to learn at the right place at the right time, and it is adopted because math path activities need high mobility of learning environment to situate students in authentic learning environments. In a U-learning environment, students learn with a PDA, WebPad, Tablet PC or laptop, in indoor, outdoor, individual, and group situations. Mobile devices and context-aware systems can sense the situation of learners, provide adaptive support to students, share and keep the process of each student’s problem solving immediately. So, the above-mentioned restrictions of traditional paper-and-pencil based math path can be overcome by generating a U-learning environment.

Although U-learning seems to be able to improve the traditional math path, only a few studies have attempted to apply this innovative approach to math paths. Most of the previous U-learning studies have been conducted on natural science courses (Chiou et al., 2010; Chu, Hwang, Huang, & Wu, 2008; Chu et al., 2010; El-Bishouty, Ogata, & Yano, 2007; Hwang et al., 2010; Peng, Chuang, Hwang, Chu, Wu, & Huang, 2009; Rogers, Price, Randell, Fraser, Weal, & Fitzpatrick, 2005), language training courses (Joiner, Nethercott, Hull, & Reid, 2006; Liu & Chu, 2010; Ogata & Yano, 2003, 2004; Zurita & Nussbaum, 2004), complex science experiments (Hwang, Yang, Tsai, & Yang, 2009), or museum learning (Hall & Bannon, 2006; Laine et al., 2009). And, mobile devices applied in these studies are almost PDAs (Chen et al., 2008; Chiou et al., 2010; Chu et al., 2008; Chu et al., 2010; El-Bishouty et al., 2007; Huang, Huang, & Hsieh, 2008; Hwang, Chu, Shih, Huang, & Tsai, 2010; Hwang, Kuo, Yin, & Chuang, 2010; Hwang et al., 2008; Hwang et al., 2009; Joiner et al., 2006; Liu & Chu, 2010; Liu et al., 2009; Ogata, Saito, Paredes, San Martin, & Yano, 2008; Ogata & Yano, 2003, 2004; Peng et al., 2009; Rogers et al., 2005; Yang, 2006; Zurita & Nussbaum, 2004) or mobile phones (Chen & Chao, 2008; Laine et al., 2009; Milrad & Spikol, 2007). Additionally, it is evident that these ubiquitous learning studies have aimed to integrate real-life situations with digital learning environment during the learning process, but lack the scheme for individual assessment and remedial instruction. In instructional practice, it is also imperative to investigate how to administer adaptive tests in the U-learning environment, as well as to provide follow-up guidance based on the individual testing reports.

Therefore, this study attempts to develop an adaptive U-learning math path system which combines U-learning, campus math path and the previously developed Knowledge structure based adaptive testing system (KSAT). This system can extend the application of U-learning to mathematics learning in elementary schools. KSAT is the improvement in computerized adaptive testing system of Diagnosys (Appleby, Samuels, and Treasure-Jones, 1997). It can provide adaptive testing functions, as well as adaptive remedial instructions to improve the shortcomings of testing and remedial education in previous studies. After establishing the adaptive U-learning math path system, the instructional effect of this system is evaluated through experiment research. With regard to the learning content of experiment analysis, only the “looking for patterns” unit in a fifth-grade math curriculum is taken as an example in this study.

To sum up, there are three major differences between this study and previous literatures. First, this study applies U-learning to a novel exercise and it hasn’t been explored in the literatures. Second, the mobile device adopted in this study is Tablet PC that is seldom used in previous applications. The reason of applying Tablet PC is that the solutions or problem solving process of math items are hard to write down by small mobile devices such as PDAs or mobile phones. Third, the proposed adaptive U-learning system provides adaptive testing functions, as well as adaptive remedial instructions to improve the shortcomings in previous studies.

The goals of this study are presented as follows:

1. To design an adaptive U-learning math path system.
2. To compare the influence of adaptive U-learning math path against traditional paper-and-pencil based math path on mathematical learning scores in the “looking for patterns” unit.
3. To compare the influence of adaptive U-learning math path against traditional paper-and-pencil based math path on remedial learning scores of mathematics in the “looking for patterns” unit.
4. To explore the effect of adaptive U-learning math path method on the learners’ mathematical connection abilities after learning the “looking for patterns” unit.
Adaptive U-learning Math Path System

System Overview

Figure 1 shows the adaptive U-learning environment with RFID sensors and wireless networks. Each target site has an RFID tag attached to it which records the identification data of the site. Each student is equipped with a Tablet PC with an RFID reader which can read the data from the tag if the student is close enough. Once the adaptive U-learning system identifies the site, relevant information can be read from the Learning Activity, Learning Portfolio, Testing Portfolio and Remedial Learning databases in the server via wireless communications (see Figure 1). The adaptive U-learning math path activities will be described in the following paragraph.

Instruction Contents

The math path activity of the “looking for patterns” unit in the adaptive U-learning math path system is divided into three main sessions based on the experts’ knowledge structure (see Figure 2). The main points of the design include the following: The content of the first session contains instructing students to predict the next time period using time patterns; the students can obtain a solution based on direct arrangement patterns and positioning patterns. The content of the second session contains the use of a series of numbers to solve application problems and making predictions through number sets. The content of the third session involves the evaluation of the relationships between the numbers. Table 1 presents an example of the second session of the math path learning activity.

![Figure 1. The framework of the adaptive U-learning math path](image)

![Figure 2. A segment of the experts’ knowledge structure](image)
**Table 1. A math path learning activity sample**

<table>
<thead>
<tr>
<th>Name of challenge</th>
<th>Authentic situation</th>
<th>Learning activity and question</th>
</tr>
</thead>
</table>
| B4 Precious Books | ![Library](image)  | Challenge No. B4-1  
This is the library and it is filled with books. When you come here, please remain quiet so as not to distract the other students. Now, let us look at the books and see if we can find any patterns. The books have been arranged in categories, the same types of books are placed together based on the code numbers on the shelves.  
① What is the number of the book located five rows below book number 002031? What is the difference in number of the neighboring books?  
② What is the number of the book six rows above book number 002027? What is the difference in number of the neighboring books?  
Can you find them? Write your solution! |

**Learning environment and teaching process**

The following environment and procedures are required to use the adaptive U-learning math path system to teach mathematics, administer adaptive tests, and conduct remedial lessons:

1. After the teacher introduces the system and the usage guidelines, the students enter the adaptive U-learning math path system, and log in using their respective accounts and passwords. The students click on the system guidelines to learn about the rules of the math path challenge. Afterward, they view the map (see Figure 3) and start using the adaptive U-learning math path.

![Figure 3. System map of the adaptive U-learning math path](image)

2. After viewing the map, the system assigns a sequence of passing challenges to each student. According the given sequence, the students enter the campus learning area carrying the Tablet PC. When the students enter the first campus area, the RFID reader in the Tablet PC can read the data from the tag in this area actively. All students should first view the online math tutorial courses (see Figure 4) and learn math concepts of the unit that are required to solve the math path challenge in this area. After viewing all courses, the students can press the challenge button on the lower end of the screen and read the rules of the math path challenge.

3. After entering the challenge, the students click the ‘problems’ button under ‘file’ on the upper corner of the screen. When the questions appear on the screen, the students can solve real situation problems using the U-learning system, and use their Tablet PCs to record their solutions and answers in the blank portion under the questions (see Figure 5). When the activity is completed, the students must press the ‘save’ button under ‘file’ to transmit their answers to the U-learning system.
4. After the students complete an activity, they click the ‘group discussion’ button on the left to check the solutions of the other students and conduct discussions in small groups (see Figure 6). When they evaluate all solutions of other groups, they must click the ‘vote’ button to choose the best solution.

5. When the students complete a session, they must return to the classroom and the teacher can explain unclear concepts and make a conclusion of the session.
6. After the students complete all the sessions, they can click the ‘test’ button to enter the KSAT computerized adaptive testing system (see Figure 7).

7. When a student completes the test, the computer screen will show an individual assessment report of his test results. The upper portion of the report includes the student’s basic information, the score and percentile of this test. The lower half includes the student’s learning records, test dates, test times, scores, the number of questions, and the recorded answers. Additionally, the report also lists details of the concepts tested in this unit, the diagnosis results, and online remedial materials of the math path (see Figure 8-9).

---

**Figure 8.** KSAT concepts diagnostic report 1

---

**Figure 9.** KSAT concepts diagnostic report 2

---

Subsequently, the respective remedial instruction is implemented using online remedial materials of the math path. The U-learning math path remedial lesson is displayed in Figure 10.

---

**Figure 10.** Animation clip of remedial lesson in the system
Participants and Experimental Procedure

The participants of this study are elementary school fifth graders in Taiwan. Four classes are chosen to participate in this study, two classes are assigned as the experimental group randomly, and the other two are assigned as the control group. There are originally 122 students in the four classes. However, 4 students are unable to complete the tests; thus, the effective sampling population is reduced to 118 students, with 59 in the experimental group and 59 in the control group. The experimental group is taught using “the adaptive U-learning math path learning activity,” while the control group learned using “the traditional paper-and-pencil based math path learning activity”. The experiment process is shown in Table 2:

<table>
<thead>
<tr>
<th>Process (time)</th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1 (one lesson)</td>
<td>computerized adaptive diagnostic test</td>
<td>computerized adaptive diagnostic test</td>
</tr>
<tr>
<td>Teaching activities (six lessons)</td>
<td>adaptive U-learning math path teaching process</td>
<td>traditional math path teaching process</td>
</tr>
<tr>
<td>Test 2 (one lesson)</td>
<td>computerized adaptive diagnostic test</td>
<td>computerized adaptive diagnostic test</td>
</tr>
<tr>
<td>Remedial activities (one lesson)</td>
<td>adaptive U-learning remedial lessons</td>
<td>traditional remedial lessons</td>
</tr>
<tr>
<td>Test 3 (one lesson)</td>
<td>computerized adaptive diagnostic test</td>
<td>computerized adaptive diagnostic test</td>
</tr>
<tr>
<td>Mathematical connection ability questionnaire (twenty minutes)</td>
<td>paper-and-pencil test</td>
<td>paper-and-pencil test</td>
</tr>
</tbody>
</table>

As shown in Table 2, there are three computerized diagnostic tests of “looking for patterns” unit and a mathematical connection ability questionnaire used as evaluation tools. The three tests are Test1, Test2 and Test3, each with 30 questions. They are parallel forms with a Cronbach α reliability coefficient of 0.75, indicating that the difficulty rates of the test questions are between 31.3% and 96.0%. The coefficient of determination ranged between 0.3 and 0.5, and the validity is assessed by field experts.

In order to combine ubiquitous learning, campus math path and the knowledge structure based adaptive testing system. Test1, Test2 and Test3 are all administered by KSAT. The adaptive testing algorithm of KSAT is based on the students’ item structure. This structure is estimated by ordering theory (Airasian & Bart, 1973). As shown in Figure 11, if the subject gets a top skill (item C) correct then it is inferred that he or she also understands its prerequisites (items F, G, H, I). This algorithm can predict students’ profiles using fewer items than in original paper-and-pencil based tests.

![Figure 11. The students’ item structure](image)

The mathematical connection ability questionnaire is developed based on mathematical connection competence indicators of Grade 1-9 Curriculum Guidelines in Taiwan (Ministry of Education of the R.O.C., 2009). There are 3 dimensions in this questionnaire: the first dimension assesses the students’ abilities to connect mathematics to real
life situations or to other subjects; the second dimension tests the students’ abilities to transfer mathematical problems in daily life into equations; the third dimension investigates the students’ abilities to solve problems in daily life using appropriate mathematical methods. Each dimension has two items, one is yes/no item and the other is open-ended item. If a student’s answer to the yes/no item is “yes”, and a right example or reason is also given in the open-ended item, it’s scored “correct” (1), otherwise it is scored “wrong” (0). According to the results of this questionnaire, students getting high scores show they have better performances in mathematical connection. The full questionnaire (three dimensions and six items) is shown in Table 3. Because the original questionnaire is designed in Chinese with the words learned by fifth grade students, students have no difficulties in reading. Before formal experiment, this questionnaire has been tried out in pilot study to make sure that fifth graders can understand these items and the Cronbach’s α reliability coefficient of questionnaire is 0.743.

<table>
<thead>
<tr>
<th>Dimension of questionnaire</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Consciousness</td>
<td>Item1-1: After learning through the math path, I can understand more about connecting mathematics to daily life or other subject areas. For example, the numbers of seats in the theater are arranged in a regular math pattern. □ yes (please give an example at least) □ no Item1-2: My example or reason:</td>
</tr>
<tr>
<td>2. Transformation</td>
<td>Item2-1: After learning through the math path, I can understand more about how to convert mathematical problems in real life in the form of mathematical symbols and formulas. For example, if the bottom surface of a sandwich is a right triangle, we name it △ABC, and ∠A = 90°, ∠B = 30°, then ∠C = 180° − 90° − 30° □ yes (please give an example at least) □ no Item2-2: My example or reason:</td>
</tr>
<tr>
<td>3. Problem-solving</td>
<td>Item3-1: After learning through the math path, I can understand more about how to use appropriate mathematical methods to solve problems in daily life. For example, we can use subtraction to solve the question “The ball is ten dollars. I gave him fifty dollars, how much change should he give me?” □ yes (please give an example at least) □ no Item3-2: My example or reason:</td>
</tr>
</tbody>
</table>

**Experiment Results**

The experiment results are reported in three parts: First, the influences of the adaptive U-learning math path and traditional paper-and-pencil based math path towards the learning achievements of students are investigated. Second, the effects of remedial learning methods towards the students’ remedial learning achievements are compared. Third, the differences in the students’ mathematical connection ability under the two different learning modules are explored.

**Teaching activities performance**

In order to investigate the influences of the adaptive U-learning math path activities towards the learning achievements of students, this study analyzes with an analysis of covariance (ANCOVA). The covariance is Test1 score, the independent variable is the teaching model (group: experimental group and control group), and the dependent variable is the Test2 score. The test of homogeneity of regression coefficients is administered before ANCOVA, obtaining the results $F = .002$ and $p = .969 > .05$; this indicates that it fulfills the basic assumption of the homogeneity of the regression coefficients in ANCOVA. Then, One-way ANCOVA is implemented. After removing
the influence of Test1 scores, an effect test is administered on the Test2 scores and the results obtained are shown in Table 4, namely $F = 4.13$, $p = .045 < .05$, indicating that there is a significant difference between two groups.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>Df</th>
<th>Mean Squares</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1002.27</td>
<td>1</td>
<td>1002.27</td>
<td>4.13</td>
<td>.045</td>
</tr>
<tr>
<td>Error</td>
<td>27928.42</td>
<td>115</td>
<td>242.86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This result shows that the Test2 scores of the students varied due to the different teaching models they are exposed to. Meanwhile, based on the adjusted mean values, the experimental group obtained 49.66 points while the control group obtained 43.82 points. The mean score of the experimental group is significantly higher than that in the control group.

**Remedial learning activities performance**

Next, we compare the effects of adaptive U-learning remedial learning methods towards the students’ remedial learning achievements. This is similarly calculated using the analysis of covariance (ANCOVA). The covariance is the Test2 score, the independent variable is the remedial learning model, and the dependent variable is the Test3 score. The test of homogeneity of regression coefficients obtains the results $F = 3.42$ and $p = .067 > .05$; this indicates that it fulfills the basic assumption of the homogeneity of the regression coefficients in ANCOVA and One-way ANCOVA is then implemented. After removing the influence of the Test2 scores on the Test3 scores, an effect test is administered on the Test3 scores and obtained the results shown in Table 5, namely $F = 4.438$ and $p = .037 < .05$, indicating that there is a significant difference between two groups.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>Df</th>
<th>Mean Squares</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>416.67</td>
<td>1</td>
<td>416.67</td>
<td>4.438</td>
<td>.037</td>
</tr>
<tr>
<td>Error</td>
<td>10797.70</td>
<td>115</td>
<td>93.89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The above-mentioned result shows that the Test3 scores of the students differ greatly between the adaptive U-learning math path remedial lessons group and the traditional paper-and-pencil based math path remedial lessons group. Meanwhile, based on the adjusted mean values, the experimental group obtained 60.02 points while the control group obtained 56.20 points. The mean score of the experimental group is significantly higher than that in the control group.

**Mathematical connection ability performance**

The differences in mathematical connection ability performances are obtained by administering a Chi-square test on the answers of the experimental group and the control group in the mathematical connection ability questionnaire. The test results are shown in Table 6.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Group</th>
<th>Correct (%)</th>
<th>Wrong (%)</th>
<th>$\chi^2$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.Consciousness</td>
<td>Experimental</td>
<td>43(72.9)</td>
<td>16(27.1)</td>
<td>4.427</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>32(54.2)</td>
<td>27(45.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.Transformation</td>
<td>Experimental</td>
<td>44(74.5)</td>
<td>15(25.5)</td>
<td>4.523</td>
<td>.033</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>33(55.9)</td>
<td>26(44.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.Problem-solving</td>
<td>Experimental</td>
<td>42(71.2)</td>
<td>17(28.8)</td>
<td>5.130</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>30(50.8)</td>
<td>29(49.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to results in Table 6, $p = .035 < .05$ in dimension1, $p = .033 < .05$ in dimension2, and $p = .024 < .05$ in dimension3 indicating that there is a significant difference between two groups in each dimension. More students in the experimental group can indicate appropriate examples that connect mathematical concepts with their life, such as
flower patterns, the bricks on the walls, locations of the map, or rate of exercise. Overall, the experimental group students perform better than the control group in all dimensions of mathematical connection ability.

**Conclusion**

In this study, we propose an adaptive U-learning system for math path activities and compare its performance to traditional paper-and-pencil based math path activities. The conclusions are as follows:

1. The adaptive U-learning system developed in this study can extend the application of U-learning to mathematics learning in elementary schools, and provide adaptive testing functions as well as adaptive remedial instructions to improve the shortcomings in previous studies.
2. After the experimental teaching sessions, the results indicate that the adaptive U-learning math path is conducive to the improvement of students’ mathematical achievements.
3. The results of the remedial lessons indicate that the adaptive U-learning math path remedial lessons are more effective than traditional paper-and-pencil based math path remedial lessons.
4. Regarding the performance of mathematical connection ability questionnaire, the experimental group performs better than the control group in dimensions of “consciousness”, “transformation” and “problem-solving”.

The above conclusions show that the proposed instructional model and learning system are better than the traditional instructional model in learning “looking for patterns” unit. However, some people may doubt of the feasibility of the proposed instructional model because the prices of Tablet PCs are high. In Taiwan, Tablet PCs are also expensive. Only a few urban elementary schools, learning centers and cram schools can afford such devices to teaching. But, we think U-learning or pervasive learning is still a promising learning model and will become popular in the future because the mobile devices with big screen such as iPAD are getting cheaper with the rapid development of technologies. Furthermore, the study still leaves room for improvement, such as applying to other units, increasing the assessment module of students’ problem-solving process. These will be explored in future research.

**Acknowledgments**

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 93-2521-S-142-004- and NSC 98-2511-S-142-003-M.

**References**


The Development of Epistemic Relativism versus Social Relativism via Online Peer Assessment, and their Relations with Epistemological Beliefs and Internet Self-efficacy

Chin-Chung Tsai
Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, 106, Taiwan // cctsai@mail.ntust.edu.tw

(Submitted November 6, 2010; Revised February 12, 2011; Accepted March 26, 2011)

ABSTRACT
Online peer assessment has been advocated by numerous contemporary educators. This study interviewed forty students who had experienced an online peer assessment activity for learning. Each of these students was asked to complete a research proposal in an educational method course for peer assessment. Using the online peer assessment system, the students, who performed the roles of both authors and reviewers, submitted their proposals, reviewed their peers’ work and obtained peers’ comments. Based on the interview data, sixty percent of the students could gradually develop views of “epistemic relativism” and “social relativism” through the online peer assessment activity. “Epistemic relativism” refers to recognition of the diversity of knowledge perspectives involved in a research issue, while “social relativism” indicates an understanding of the multiplicity of peer perspectives. These two views, which are perceived as positive impacts of online peer assessment, have rarely been documented in the relevant literature. This study further found that students’ epistemological beliefs were related to their views of “epistemic relativism,” while their Internet self-efficacy was associated with their views of “social relativism.” More sophisticated epistemological beliefs and higher Internet self-efficacy can likely facilitate the development of views regarding “epistemic relativism” and “social relativism.”

Keywords
Online peer assessment, relativist, Internet self-efficacy, epistemological beliefs

Introduction
Online peer assessment (PA) has been advocated by numerous contemporary educators (Cathey, 2007; Chen & Tsai, 2009; Sithiworachart & Joy, 2008). Online PA for learning requires the participating students to complete a learning task or assignment for peers to make evaluations of and give comments on, such as completing a research proposal (Wen & Tsai, 2008), or designing a detailed plan or activity (Tsai & Liang, 2009; Tseng & Tsai, 2007). The research literature has documented that PA in general, or online PA in particular, can improve the participants’ understandings in the cognitive and metacognitive domains, and enhance their social and thinking skills (Topping, 1998; Tsai, Lin & Yuan, 2002).

However, this study attempted to explore some other benefits of using online PA. Research (e.g., Yang, Y. F. & Tsai, 2010) has indicated that online PA can help participants acquire a variety of perspectives, thus developing so-called “relativist” views. The relativist views in general support the diverse perspectives of other theories and other peers (Tsai, 2004). This study further differentiated two forms of relativist views, including “epistemic relativism” and “social relativism.” “Epistemic relativism” refers to recognition of the diversity of knowledge perspectives involved in the online PA project, while “social relativism” indicates an understanding of the multiplicity of peer perspectives. “Epistemic relativism” asserts that the validity of each theoretical perspective is only relative to each other, whereas “social relativism” acknowledges the diversity of peers’ ideas. Relatively speaking, the latter is likely more akin to the position of multiplism (Perry, 1970) while the former is a more developed critical stance regarding knowledge claims. In light of this interpretation based on Perry’s study, “epistemic relativism” may be a more sophisticated position than “social relativism.” In this paper, these two views were discussed by sharing some interview data derived from a PA study with higher education students. By doing this, additional strengths of implementing online PA were illustrated.

Moreover, this study explored two factors which may be related to the occurrence of these relativist views. The first factor is the students’ epistemological beliefs, that is, their beliefs about the nature of knowledge and knowing (Hofer, 2001, 2010; Wong & Chai, 2010). Educators have concluded that learners’ epistemological beliefs are associated with their learning strategies, reasoning modes and knowledge acquisition when processing or acquiring information.
(Hofer & Pintrich, 1997; Liu, Lin & Tsai, 2011; Tsai, 1998). The findings from previous studies have revealed that students holding more sophisticated epistemological beliefs (highlighting the uncertainty, multiple sources of knowledge) tend to adopt better cognitive approaches and attain higher learning outcomes than those possessing more shallow views about the nature of knowledge and learning, such as emphasizing the absolute status of knowledge and the role of memorization in learning (Greene, Muis, & Pieschl, 2010; Tu, Shih, & Tsai, 2008). A recent study completed by Tsai and Liang (2009) revealed that students with more advanced epistemological beliefs tended to greatly improve their original work and provide more high-quality comments to their peers in online PA learning environments. Hence, it is expected that students’ epistemological beliefs are related to their views of and gains in online PA.

The second factor considered in this study is the Internet self-efficacy possessed by the students. “Self-efficacy” refers to an individual’s beliefs and expectations in his/her ability to perform a task, and it affects the individual’s processing of the task, ways of completing it, efforts to be devoted, and how he or she will maintain effort in handling demanding situations (Bandura, 1977, 1996). Similarly, Internet self-efficacy indicates students’ self-perceived confidence and expectations of using the Internet (Liang & Tsai, 2008; Tsai & Tsai, 2010). It has been proposed that learners with higher Internet self-efficacy may have better attitudes toward Internet and Internet-related learning activities, and they would have a greater chance of success in computer and Internet-related tasks (Peng, Tsai & Wu, 2006; Tsai & Tsai, 2003; Wu & Tsai, 2006). Consequently, it is believed that students’ Internet self-efficacy may be associated with their views and gains derived from online PA learning environments.

In sum, this study, through gathering a group of students who experienced online peer assessment, attempted to explore:
1. Whether the students expressed relativist-oriented views when reflecting their experiences of online PA.
2. How the students’ relativist views were associated with their epistemological beliefs and Internet self-efficacy.

**Method**

**Participants**

This study was conducted at a research-oriented university in North Taiwan, and included forty-five graduate students with a major in education. All of them were enrolled in an educational research method course. As required by the course, each of the students should complete a research proposal, which should be submitted to an online system for peer-review. All of them had relevant computer or Internet abilities to complete the online learning task.

**Online PA learning activity**

After eight weeks of the course, each of the students developed an initial educational research proposal, and submitted it to an online system. All proposals experienced three-round online PA treatment, and the assessment process was undertaken in an anonymous way. The online peer assessment procedure was based on a previous online PA model proposed by Tsai, Lin and Yuan (2002) and Tsai, Liu, Lin and Yuan (2001). Each proposal needed to be refined and submitted three times. The process of the online PA activity is as follows: The students submitted their original proposals; they reviewed their peers’ work; they submitted their revised proposals; they reviewed their peers’ work again, and finally, they submitted their final proposals, and completed the final peer reviews. Each participant acted both as an author and a peer reviewer. For each round of PA, each participant reviewed about five proposals drafted by their peers. The participants reviewed the same proposals assigned across different rounds of peer assessment. The on-line PA took about eight weeks.

**Data Collection and Analysis**

Interview. Each participant in this study was interviewed individually by a trained researcher after finishing the three-round online PA. As five students, for some unexpected reasons, could not contribute to the interviews, this study conducted individual interviews for a total of forty students. The interview questions mainly focused on the
experiences of taking part in the online PA learning activity, and their views as well as perceived gains from learning via online PA. In addition, the students’ perceptions of their peers’ comments were explored. Sample interview questions were: In any aspect, how did you learn from the online PA? What did you gain from the online PA? How did you perceive the peer comments from the online PA? What did you think about and how did you react to your peers’ comments?

All of the interviews were audio-recorded. The interviews were conducted in Chinese and then fully transcribed for further analysis. The interview quotations presented later in this paper were those perceived as being the most representative or the most fruitful ideas expressed by the interviewed students. One researcher coded each student’s interview responses to examine the possibility of showing relativist-oriented views. One additional independent researcher, who actually read all of the interview transcripts, validated the coding. The agreement of both researchers was around 0.90. The responses with disagreement were resolved upon discussion.

The assessment of epistemological beliefs. To assess the students’ epistemological beliefs, this study utilized the questionnaire developed by Chan and Sachs (2001). The questionnaire, exploring students’ epistemological beliefs about learning, included nine items, each with three options: two corresponding to a naïve, shallow view about learning, while one reflected a deep and more sophisticated view. A Chinese version of this questionnaire had been used in another study (Tu et al., 2008). The following is a sample item from the questionnaire.

The most important thing you can do when you are trying to learn science is
a. faithfully do the work the teacher tells you to do.
b. try to see how the explanation makes sense.
c. try to remember everything you are supposed to know.

Students’ responses were scored 1 point if the answers responded to a more advanced position (e.g., the “b” option in the sample item), while those reflecting a naïve view of learning were given 0 points (e.g., the “a” and “c” options in the sample item). The reliability coefficient (KR20) was estimated around 0.68 for the nine items. Although the coefficient is not very high, it is still considered as satisfactory. Through using the questionnaire, this study acquired an Epistemological Belief Score (EBS) for each student (ranging from 0 to 9), with a higher score on the questionnaire indicating stronger agreement with the more advanced epistemological beliefs.

The measurement of Internet self-efficacy. The Internet Self-efficacy Survey (ISS) employed in this study was adapted from original items developed in previous studies (Peng et al., 2006; Wu & Tsai, 2006). These studies have already proved the adequate validity and reliability of the ISS. The ISS included nine items. The items were presented with bipolar strongly confident/ strongly unconfident statements in a seven-point Likert mode. Sample items are: “I am good at searching for information on the Internet,” and “I think I can talk to others in online chatrooms.” Each student obtained an average score from the ISS items (ranging from 1 to 7), with a higher score indicating higher Internet self-efficacy.

Results

Epistemic relativism

First, it was found that the use of online PA can enhance students’ epistemic awareness, acknowledging the relativist view about different theories. The relativist position asserts that there is no certainly right or wrong knowledge; rather, there are multiple interpretations of any issue studied. The validity of each theoretical perspective is only relative to each other (Perry, 1970). This view, called “epistemic relativism,” was acknowledged by many students after experiencing the online PA in this study. For example, they gave the following responses during the individual interviews:

- Through online peer assessment, the peer comments helped me find that knowledge or theories in various fields may be related to my work.
- In the beginning, I thought the proposal I was working on was quite simple and straightforward. There was even no ambiguity about it. However, when I submitted it for online peer assessment, I surprisingly found that there were still numerous theoretical perspectives raised for debate.
Through online peer assessment, I realized that I just think about one side of the conceptual viewpoint, and totally ignore the others. It is clear that through online PA, the students gathered multiple/different knowledge perspectives of their work. However, they realized that each of these perspectives, though not certainly right or wrong, was relatively differently valid. Hence, critical thinking and careful judgment of peers’ comments are quite important in the online PA process. For example, the students responded that:

- Different suggestions came from different peers. I gathered all the conceptual perspectives to be evaluated and finally improved my own proposal.
- By online peer assessment, my peers evaluated my work. But, when I got their evaluations, interestingly, I needed to carefully “evaluate” their evaluations. Then, I could know which one was more applicable.
- I learned a lesson from the online peer assessment; that is, do not treat everyone’s comments as equally important. I needed to think over and over again about every peer comment.

Based on these responses, the students learned that not all of the comments were equally important. The participating students should have relative weights for considering peers’ comments. The acquisition of multiple knowledge perspectives, and careful reflection on the relative importance as well as validity of these perspectives constitute the main ideas about “epistemic relativism” for online PA.

Social relativism

The other view derived from the online PA is the social acceptance of peers’ opinions, personal preferences and subjective comments. This is called “social relativism,” and recognizes that everyone has personal ideas concerning an issue, and everyone expresses ideas based on his/her preference according to the relative connection to the context of the issue concerned. By implementing online PA, it was gradually found that the students tended to socially accept the diversity of their peers’ opinions. For example, the students responded that:

- I tried to make sense of what kind of peer comment was conveyed. Finally, I realized that not everyone has the same point of view.
- By reading the peer comments online, I learned that I need to accept that everyone has his/her own opinions.
- Through online peer assessment, I developed an appreciation of and respect for the diversity of peer comments. I think this may be quite useful for my future career. (Researcher: How?) In the future, I will also face quite different opinions from peers in the workplace.
- After experiencing the online PA, I suddenly found that I had become more open-minded to any piece of peer feedback.

It is clear that the students gained a better viewpoint from the variety of peer comments. In addition, in terms of the negative or unfair peer evaluations, the students gradually developed better attitudes toward and adaptations of them. For example, they stated that:

- To be honest, I once felt very frustrated about the online peer assessment. I found some comments were very useful, but still many of them were just personal preferences. But, now, I think this is just “individual difference”….. This is the real world.
- Online peer assessment has helped me to develop an adequate acceptance of negative comments. I need to adjust myself to these negative comments….. As the online peer assessment is anonymous, they are not so harmful.

In sum, by utilizing online PA, the students gained the social recognition of other opinions or others’ individual preferences.

The distribution of epistemic relativism and social relativism

Table 1 shows the distribution of the participants who expressed ideas of “epistemic relativism” and “social relativism” during the individual interviews. As aforementioned, two researchers coded each student’s interview responses, and examined if his/her responses demonstrated “epistemic relativist” and/or “social relativist” views. The coding results are presented in Table 1.
Table 1. The number of interviewed students showing the ideas of epistemic and social relativism

<table>
<thead>
<tr>
<th></th>
<th>Social relativism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not shown</td>
</tr>
<tr>
<td>Epistemic relativism</td>
<td></td>
</tr>
<tr>
<td>Not shown</td>
<td>16</td>
</tr>
<tr>
<td>Shown</td>
<td>4</td>
</tr>
</tbody>
</table>

Based on the results in Table 1, eleven among the forty interviewed students stated ideas of both “epistemic relativism” and “social relativism.” Four displayed “epistemic relativism” only and nine displayed “social relativism” only. More students had ideas of “social relativism” (n = 20) than of “epistemic relativism” (n = 15). Forty percent of the students (n = 16) did not mention any idea related to “epistemic relativism” or “social relativism.”

The role of epistemological beliefs

The interviewed students were divided into advanced epistemological beliefs and naïve beliefs according to their mean scores assessed by the Epistemological Belief questionnaire (mean = 4.37 for the total scores from 0 to 9). Consequently, 19 students were labeled as holding advanced epistemological beliefs (score 5-9 for the nine-item questionnaire), while the rest were classified as having naïve epistemological beliefs. The interplay between student epistemological beliefs and their views of “epistemic relativism” and “social relativism” is presented in Table 2.

Table 2. The role of epistemological beliefs in the views of epistemic relativism and social relativism

<table>
<thead>
<tr>
<th></th>
<th>Epistemic relativism (n)</th>
<th>Social relativism (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced epistemological beliefs (n = 19)</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Naïve epistemological beliefs (n = 21)</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

Based on the results in Table 2, by and large, the students with more sophisticated epistemological beliefs tended to have views of “epistemic relativism” and “social relativism.” About a half of the students in the advanced epistemological belief group expressed ideas of “epistemic relativism” and “social relativism” (n = 11, 11 respectively). Still, approximately a half of the students in the naïve epistemological belief group possessed views of “social relativism” (n = 9). However, only less than a quarter of the students in the naïve epistemological belief group (n = 4) held views of “epistemic relativism.” Therefore, it is concluded that epistemological beliefs may play a more important role in “epistemic relativism” than in “social relativism.” More sophisticated epistemological beliefs can help the development of “epistemic relativism” views for those students who experience online PA.

The role of Internet self-efficacy

Similarly, the students were divided into high and low Internet self-efficacy groups by their mean scores on the ISS questionnaire (mean = 5.56 for the Likert scale of 1-7). Consequently, 20 students were labeled as having high Internet self-efficacy while 20 were categorized as having low Internet self-efficacy. The relationship between Internet self-efficacy and views of “epistemic relativism” and “social relativism” held by the students is presented in Table 3.

Table 3: The role of Internet self-efficacy in the views of epistemic relativism and social relativism

<table>
<thead>
<tr>
<th></th>
<th>Epistemic relativism (n)</th>
<th>Social relativism (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Internet self-efficacy</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Low Internet self-efficacy</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

According to Table 3, almost the same number of students in the different levels of Internet self-efficacy expressed views of “epistemic relativism” (n = 8 and 7, respectively). Nevertheless, many more students in the high Internet self-efficacy group displayed ideas of “social relativism” than those in the low Internet self-efficacy group (n = 13 versus n = 7). Thus, Internet self-efficacy seems to have an effect on “social relativism,” but not on “epistemic
relativism.” High Internet self-efficacy tends to facilitate the development of views regarding “social relativism.”

Discussion

This study identified the views of “epistemic relativism” and “social relativism” expressed by the students who participated in the online PA activity. To develop views of “epistemic relativism” and “social relativism” by PA, the online technology plays an important role in achieving this. First, it ensures better anonymity for PA (Tsai, 2009). The anonymous PA environment assisted by online technology shapes a relatively neutral learning context in which peers can solely focus on evaluating the work, not on the person who did the work. As shown by one student earlier (i.e., the negative comments are not so harmful), this is especially useful for the development of “social relativism” because of the better anonymity and de-contextualized peers in the online environment (Tsai, 2001a). Second, by way of the online PA system, in a short period of time, the participants can gather a variety of peer comments without the constraints of time and location. Also, many “outside” peers can be easily invited to judge the work, and a variety of different knowledge perspectives and opinions can emerge. By exposure to such diversity, students’ views of “epistemic relativism” and “social relativism” can be potentially triggered.

This study further suggests that sophisticated epistemological beliefs can help the development of “epistemic relativism” for those students engaged in online PA. Since the students with more mature epistemological beliefs tend to perceive knowledge as uncertain and coming from multiple sources (Hofer & Pintrich, 1997), they are likely to express the acknowledgement of various theoretical perspectives, the view of “epistemic relativism” proposed in this paper. Moreover, the analysis between Internet-self-efficacy and their views of social relativism indicated that high Internet self-efficacy seems to facilitate the development of “social relativism” during learning via online PA. Past research has indicated that high Internet self-efficacy mainly comes from the users’ rich usage experiences and behaviors (Durndell & Haag, 2002; Peng et al., 2006; Wu & Tsai, 2006). These rich experiences may help the users recognize the diversity of viewpoints expressed by others in the world of the Internet. Thus, they are more likely to acquire the recognition of “social relativism” as defined in this study. The aforementioned findings have also provided evidence that “epistemic relativism” and “social relativism,” though related to each other (Table 1), are still different, as they are associated with different factors (i.e., epistemological beliefs, Internet self-efficacy). These findings are also consistent with the perspective discussed previously that “epistemic relativism” may be more advanced than “social relativism.” That is, “epistemic relativism” is more related to some underlying philosophical positions or higher-order thinking (such as epistemological beliefs, Tsai, 2001b; Yang, F.-Y. & Tsai, 2010) whereas “social relativism” is more associated with the students’ perceived skill of using the Internet (i.e., Internet self-efficacy), considered as a more behavior-oriented and relatively less profound factor. However, this study provides evidence that “epistemic relativism” and “social relativism” may have reciprocal interactions with each other.

Concluding remarks

This paper presents empirical support that many students who participate in online PA could develop views of “epistemic relativism” and “social relativism.” The diversity of ideas among peers cannot guarantee the achievement of “epistemic relativism” and “social relativism,” but it is their prerequisite. Online PA provides a satisfactory learning environment for evoking the diversity of different perspectives. Then, the teacher is not the sole knowledge source for learning. The view of “relativism” is quite sophisticated in terms of epistemological development (Perry, 1970). The findings of this paper also concur with the assertion proposed by researchers (Mason & Boldrin, 2008; Mason, Boldrin & Ariasi, 2010; Tsai, 2004) that the Internet should not be regarded simply as a cognitive or metacognitive tool; rather, it should be utilized as an “epistemological” tool. The proper use of online learning environments can promote epistemological development for students. It is recommended that researchers interested in online PA elaborate the ideas of “epistemic relativism” and “social relativism” proposed in this paper to observe students’ gains in online PA in a broader sense. Teachers who implement online peer assessment for learning may also monitor students’ possible acquisition of these relativist views. If possible, the teachers may have explicit discussions of these views during the implementation of online peer assessment. By such discussion, either face-to-face or online, these views may be fostered.

This study also investigated the factors related to the views of “epistemic relativism” and “social relativism.” Both
epistemological beliefs and Internet self-efficacy were identified as potential factors. If researchers agree that the acquisition of the views of “epistemic relativism” and “social relativism” is one of the important goals for learning via online PA, more studies are necessary to explore other factors that may be related to these views.

This study had certain limitations. First, it employed only a small sample size. A larger-sample study is recommended. Also, this study used questionnaires to assess students’ epistemological beliefs and Internet self-efficacy. Other methods, such as interviews or observations may be needed for future research.

Future research should be conducted to carefully examine the views of “epistemic relativism” and “social relativism” during online peer assessment. For example, educational researchers can conduct in-depth analyses to appraise if these views are involved in peer feedback or the revision of students’ original work. How to promote these views by other instructional strategies such as online argumentation or inquiry may also be a potential research issue for educators. More long-term study or instructional treatment may be required to document students’ possible progression toward “epistemic relativism” and “social relativism.”

Acknowledgments

Funding of this research work is supported by National Science Council, Taiwan, by grant numbers 99-2511-S-011-005-MY3 and 98-2511-S-011-005-MY3.

References


Yang, Y.-F., & Tsai, C.-C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction, 20*(1), 72-83.

Influence of Game Quests on Pupils’ Enjoyment and Goal-pursuing in Math Learning

Zhi-Hong Chen¹², Calvin C. Y. Liao¹, Hercy N. H. Cheng¹, Charles Y. C. Yeh¹ and Tak-Wai Chan¹

¹Graduate Institute of Network Learning Technology, National Central University, Taiwan // ²Department of Information Communication, Yuan Ze University, Taiwan // {hon, calvin, hercy, charles, chan}@cl.ncu.edu.tw

(Submitted December 23, 2010; Accepted February 15, 2011)

ABSTRACT
As a medium for learning, digital games provide promising possibilities to motivate and engage students in subject learning. In this study, a game-based learning system, My-Pet-My-Quest, is developed to support pupils’ math learning. This is due to the fact that most students in Taiwan have relatively lower positive attitude towards math learning, even though their math performance is prominent. To this end, a three-tire framework is proposed to guide the design of the My-Pet-My-Quest system. A quasi-experiment was conducted to examine the influence of game quests on pupils’ enjoyment and goal-pursuing in math learning. The results revealed that game quests were favored by students in terms of enjoyment, goal orientation, and goal intensity. Possible reasons for these results and a discussion of related issues are presented in this paper.

Keywords
Game-based learning, Game Quest, Goal-pursuing, Math learning

Introduction

Recently, game-based learning is regarded as a promising vehicle for facilitating students’ active participation and engaged learning. On one hand, this is because digital games own a number of characteristics to engage students, such as imaginary, challenges, competition, fantasy, curiosity, uncertainty (Lo, Ji, Syu, You, & Chen, 2008), goal, decision, discussion, and emotional connection (Prensky, 2008). A well-incorporation of some of these characteristics might make a boring learning be interesting and joyful. On the other hand, different learning gains have been reported, including visual attention and spatial abilities (Barlett, Anderson, & Swing, 2009), problem-solving skill (Kiili, 2007), analogical reasoning (Williams, Yuxin, Feist, Richard, & Prejean, 2007), and subject matter mastery (Ricci, Salas, & Cannon-Bowers, 1996). These positive gains further foster the advocate of game-based learning.

Based on the two reasons, some work attempts to harness these game characteristics to design a learning environment that invites students to participate in learning tasks in a more enjoyable way. Two categories of approaches are emphasized: game framework and blending approach. The former makes good use of specific game genres to develop an architecture for game-based learning. This is due to the fact that some game genres involve game characteristics that could promote students’ learning. For example, massively multiplayer online role-playing game (MMORPG) often involves quests and story-telling, which might be helpful to offer students a goal-oriented immersive learning environment (Dickey, 2007). In addition, the structure of adventure games could be helpful to students’ hierarchical knowledge learning (Hu, 2008).

In addition to these game frameworks, a blending approach is also emphasized to integrate learning activities and digital games (Gunter, Kenny, & Vick, 2008) as a game-based learning environment, in which students apply learned knowledge for game progress. A typical example is WEST system (Burton & Brown, 1979), which embeds math calculation into a board game. More specifically, students compete against computers for reaching a specific destination to win the game by combining three numbers and two operators as forward steps. Moreover, several works also try to embed learning activities into different game genres, such as a collaboration game in the Prime Climb (Manske & Conati, 2005), and a Bingo game in the EduBingo (Chang, Ching, Cheng, Chang, Chen, Wu, & Chan, 2009).
Since game-based learning has been regarded as a potential way to motivate students to learn and different approaches have also been proposed in the aforementioned studies, it seems that classroom subject learning could be enhanced through game-based learning. In this paper, we emphasize primary education and choose math as the first subject to investigate. This is due to the fact that most students in Taiwan have relatively lower positive attitude (e.g., students’ affect, value, and self-confidence) towards math learning, even though their math performance is prominent (Mullis, Martin, & Foy, 2008). In addition, if game-based learning could be applied to math learning successfully, game-based learning could also be applied to other subjects. In other words, our concern is not only how to use digital games to engage students in a joyful math learning experience, but also the flexibility of applying such learning environment to other subject learning.

To this end, the blending approach is used to develop a three-tier framework, in which domain-independent learning activities could be integrated with digital games. The intention of this framework is to maintain flexibility in dealing with different learning subjects. Moreover, based on this framework, we design a My-Pet-My-Quest system, which integrates a pet-nurturing game and a quest-delivery mechanism for math learning. This is because previous work has suggested that pet-nurturing games could be used to facilitate human-computer interaction for learning purpose (Chen, Deng, Chou, & Chan, 2007; Chen, Liao, Chien, & Chan, 2011). In addition, quest-delivery mechanism is regarded as a useful way to guide students to do tasks in role-playing games. We anticipate game quest also could be applied to subject learning, especially for increased perception of enjoyment and intention of goal-pursuing. More specifically, in this study, we focus on investigating the possibility and influences of quests on students’ math learning through the My-Pet-My-Quest system. The research question of this paper is: What are the influences of game quests on students’ math learning in terms of perception of enjoyment and goal-pursuing. We begin this paper by describing a three-tier framework, and then the development of a game environment, My-Pet-My-Quest, is presented next. Next, we proceed to discuss an empirical study on the influences of the game quest mechanism. Finally, some discussion and conclusion are drawn.

Three-tier framework

A three-tier framework proposed in this study is underpinned by a domain-independent approach. A major reason to adopt the domain-independent approach lies in the suggestion that two kinds of experts are required in the design of game-based learning (Norman, 1993), and the structure might help two kinds of experts work together. Game designers are the first experts who are good at advanced technology and know how to motivate students to learn something; the second one is domain experts who are skilled in pedagogies and know how to teach students well. A well-defined architecture or framework could be helpful to their communication and collaboration. As shown in Figure 1, the three-tier framework contains three tiers: learning activities (top tier), game world (bottom tier), and coupling mechanism (middle tier).
In terms of **game world** (bottom tier), the purpose of this tier is to stimulate and sustain students’ participation motivation by providing them with the freedom of where to go, as well as the control over what to do. The design rationale behind is underpinned by the study of virtual worlds which offers fascinating opportunities to engage students in a non-traditional and large-scale environment (Feldon & Kafai, 2008). However, different game genres have varied game world settings. Some game genres, such as shooting game or fighting game, seldom focus on the existence of a game world, whereas the design of a game world matters in other games, such as role-playing game or adventure game. Therefore, the selection of game genres should concern the settings of game world, especially its impact on students’ motivation stimulation and sustenance.

In terms of **learning activities** (top tier), the purpose of this tier is to embed different learning activities into a game world in a flexible but robust way. More specifically, the three-tier framework attempts to provide a domain-independent approach to blend a game environment with different learning activities, which could be easily designed as tasks with its pedagogies and learning goals, and then embedded into the game world. As mentioned above, such an approach might help facilitate the collaboration between game designers and domain experts, because domain experts could only focus on the design of learning activities, whereas game designers merely emphasize the development of the game environment.

In terms of **coupling mechanism** (middle tier), the purpose of this tier is to offer a goal-oriented mechanism to guide students to undertake learning tasks. This is due to the fact that a virtual learning environment could be a large-scale one, in which students might spend much time on basic system functions or task-unrelated works. In other words, to increase students’ time on-task, it is required to develop some guiding mechanisms to anchor learning tasks for students. In a sense, the coupling mechanism could be regarded as a bridge that connects and facilitates the top tier and bottom tier. Learning activities (i.e., the top tier) could be embedded into the learning tasks to undertake by students through the coupling mechanism (i.e., the middle tier), and then appear in the game world (i.e., the bottom tier). The three tiers and their descriptions are summarized in Table 1.

<table>
<thead>
<tr>
<th>Tiers</th>
<th>Purposes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning activities (Top tier)</td>
<td>To embed learning activities in different subject matters (e.g., language, math, and science learning) into the game world.</td>
</tr>
<tr>
<td>Coupling mechanism (Middle tier)</td>
<td>To provide students with a goal-oriented mechanism for guiding them to undertake learning tasks, increasing time on-task.</td>
</tr>
<tr>
<td>Game world (Bottom tier)</td>
<td>To stimulate and sustain students’ participation motivation in learning activities through a virtual game world.</td>
</tr>
</tbody>
</table>

**My-Pet-My-Quest System**

A My-Pet-My-Quest system targeting elementary students as primary users is developed according to the three-tier framework. In particular, the My-Pet-My-Quest system embeds math learning into a pet-nurturing game, allowing for an investigation of the influence of the three-tier framework.

**Pet-nurturing game (bottom tier)**

The genre of pet-nurturing game is chosen as the game world setting in this study. This is because the following two reasons. The first reason is about the *stimulation* of students’ motivation. Previous research has indicated that people, especially young students, have natural emotional attachments to their pets because the pets are simple, cute, and straightforward (Melson, 2001). Although what they mean are real pets, such emotional attachment also exists in the interaction with virtual pets (Kusahara, 2001). A typical example is Tamagotchi (Webster, 1998; Pesce, 2000), which are virtual pets that beep in several hours to ask students to feed them. When noticing the sound, the students can feed them virtual foods by pressing buttons. Although this interaction is quite simple, many children stay with them for a long period of time. This phenomenon implies that pet-nurturing might be helpful to stimulate students’ motivation and facilitate interaction. This is the reason why we choose a pet-nurturing game as the game world.
The second reason is about the sustenance of students’ participation. A pet-nurturing game often involves an economic process, needs-consumption-work. To take good care of the pets, students need to frequently check their needs. To satisfy their pets’ needs, a number of pet food and goods are purchased (i.e., consumption). Thus, students are easily encouraged to do something, including learning tasks, for earning their pets’ happy lives. Under such a structure, undertaking learning tasks could be regarded as “work.” This metaphor is similar with the working model in our work-centered society (Ransome, 2005). When learning activities could be embedded into the economic process, it might be helpful to maintain students’ participation motivation.

As depicted in Figure 2, the My-Pet-My-Quest system is developed, in which each student controls a master avatar who keeps a pet, My-Pet; thus, a major goal is to maintain the status of the pet. Each My-Pet has attributes, such as energy and mood, which are reflected in its status and can be improved through feeding and playing with My-Pet, respectively. In other words, when the student feeds the My-Pet, the “energy” attribute would be increased. Likewise, when the student plays with the My-Pet, the “mood” attribute would be increased as well. A student is therefore able to monitor achievement through the state of their pet, both in terms of happiness and healthiness.

**Learning activities (top tier)**

The learning tasks that were implemented in My-Pet-My-Quest are according to the criteria of the national curriculum for third-grade elementary school mathematics. Each unit contains three types of learning tasks: conceptual understanding, computational fluency, and problem solving. Conceptual understanding and problem solving are presented in the form of page-turners. Students are presented examples and guidelines and are asked to respond. Figure 3(a) depicts a conceptual understanding task from the unit titled *Fractions*. In this example, students fill in the answers according to the number of cups in the illustration. Similarly, Figure 3(b) presents an example of problem solving from the unit titled *Multiplication*, in which students are asked to solve a combination word problem that requires both multiplication and addition operations.
Tasks for computational fluency take the form of mini-games and are used to consolidate students’ computation skills in terms of both speed and accuracy. Figure 4 illustrates two mini-games. The mini-game on the left involves the addition of two three-digit numbers, while the mini-game on the right involves the multiplication of a two-digit number multiplied by a one-digit number. Solving these types of problems has been demonstrated to be beneficial for mathematics learning (Kilpatrick & Swafford, 2002).

**Quest mechanism (middle tier)**

Game quest is selected as a guiding mechanism in the middle tier because of the following two advantages that it brings. The first advantage is to offer a goal-oriented mechanism to guide students to participate in learning activities within a game world. Quest-taking is a graceful way in role-playing games to lead students to conduct specific tasks. When students conduct the tasks they received from non-player characters (NPCs), they might feel that they have more controls and responsibility for the tasks, although the content of the tasks is actually the same. In other words, quest-taking might turn students’ passive learning into active participation, because it offers students a clear and immediate goal to pursue. Thus, we use quest mechanism as a bridge between the bottom and top tiers.

The second advantage is to enhance the level of students’ persistence by social commitments. The quests are often received from NPCs in the game world and it could, to some extent, be regarded as a process of negotiation between students and NPCs. Therefore, we argue that receiving quests from NPCs engenders a form of social commitment, which could enhance the intensity of goal-pursuing, especially in the sub-dimensions of effort and persistence, because it is significantly related to keeping ones promises and expectancies of success. Moreover, attributions of success (Weiner, 1985; 1992) and belief in ones efforts (Dweck, 2000) have substantial influences on learning, with the implication that efforts to learning and attitudes towards persistence have benefits due to the enhancement of convictions to do something. Thus, the social commitment to NPCs might be helpful of students, both for goal-setting and for task-driven intensity.

In My-Pet-My-Quest, a quest consists of three elements: objectives, learning tasks, and rewards. An objective offers students a clear goal; a learning task is the educational content of the quest; if students complete the learning task successfully, they gain rewards. For solving math problems, for example, a quest may involve an award of EduCoins, which are only awarded to those students who can master an arithmetic task by achieving more than 90% correct responses. In other words, when students successfully complete the quests, they are rewarded with EduCoins, which can be used to buy resources for the pets, such as pet food and goods.

Quests are conducted in a number of locales, such as a store, a forest, a tower, a temple, and an arena. The flow of game quests delivered by NPCs in My-Pet-My-Quest is described as follows. First, an informer NPC appears at the door to inform the student of new quests. The student can talk to the informer to obtain more detailed information, particularly the location on the island where a quest NPC waits to escort him to the quest, as depicted in Figure 5. A student clicks the button above an NPC, and the NPC tells him the content of the quest, including objective, learning tasks, and rewards. After receiving the instructions, the student commences the quest.
Methods

This study focuses on the research question: how are the influences of game quests on students’ subject learning in terms of perception of enjoyment and goal-pursuing. To answer the question, a within-subjects experiment was conducted in an elementary school in Taiwan.

Participants

The participants in this experiment were 53 elementary students from two fourth-grade classes (median age: ten years old). In Class A there were 28 students, including 15 males and 13 females, while in Class B there were 25 students, including 14 boys and 11 girls. Since the elementary school has a policy of normal distribution at the start of the first, third, and fifth school year. Accordingly, it was assumed that the participations in each of the two groups have similar learning backgrounds and learning abilities, and that order effects have been controlled.

Instruments

Two system versions

Two different versions of the My-Pet-My-Quest system were prepared for the experiment: one with quests (i.e., via-version) and the other without quests (i.e., without-version). The two versions contain the same learning materials: two units of math activities in the mini-games about mastering the addition of two three-digit numbers. However, they are delivered to students by different ways.

With regard to the via-version, learning materials were delivered to students via game quests, which consisted of informer NPC, quest NPC, locales, learning tasks, narrative, and rewards. More specifically, learning materials in the via-version were integrated with the narratives of the quests delivered by an informer-NPC and quest-NPCs. In addition, the narratives were developed as the theme of mathematician’s problems. Some famous mathematicians, such as Newton, Euler, Gauss, and Archimedes, are calling for help about math problems. Consequently, students in the via-version would meet the informer-NPC with the message that some famous mathematicians’ math problems. Then students could explore different locales to obtain further information from certain quest-NPCs according to the cues provided by the informer-NPC. When they successfully completed learning materials, they would be rewarded by coins for their pets.

With regard to the without-version, the same learning materials were delivered by clicking a system menu, without informer-NPC, quest-NPC, and narrative, and locales. When they successfully completed learning materials, they were also rewarded by coins for their pets.
Goal-pursuing questionnaire

A goal-pursuing questionnaire developed by the first author was used to measure students’ perception. This is because the design of quests involves students’ motivation and goal-pursuing, there is no existing questionnaire available to completely reflect this aspect. Thus, we decide to develop a 5-scaled Likert questionnaire, which consisted of statements in three categories, including enjoyment, goal orientation, and goal intensity.

With regard to enjoyment, eight items were included. The sample items are “I forget about time passing while playing the game”, “I feel emotionally involved in the game”, and “I have a strong sense of achievement when completing the tasks”. They were modified from a motivational gaming scale (Dempsey, Lucassen, Haynes, & Casey, 1997) and a scale to measure students’ enjoyment of e-learning games (Fu, Su, & Yu, 2009). With regard to goal orientation and goal intensity, six and four items were contained, respectively. The two categories of items were designed according to the interpretation of goal setting by Rand (1967), involving the degrees of specificity and difficulty as well as motivational aspects, such as direction, effort, and persistence (Lock, Shaw, Saari, & Latham, 1981). The sample items are “goals were presented clearly in the beginning”, “I could establish specific goals”, and “I am stimulated to pursue certain goals.”

To increase the validity of the questionnaire, these statements were further modified by a pre-service elementary school teacher, but the reliability of the questionnaire should be clarified in the future. The purpose of the questionnaire was to understand these participants’ perception of their system use. Although the validity and reliability of the questionnaire have much room to be improved, the questionnaire could offer preliminary and immediate feedbacks.

Procedures

Since this experiment is a within-subjects design, all students were organized to experience the two system versions. To reduce the bias of treatment order, the within-subject groups were presented the two versions in a different order, as shown in Figure 6. More specifically, in Class A, participants first used the without-version for two sessions and then used the via-version for two sessions. In contrast, in Class B, participants first used the via-version for two sessions and then used the without-version for two sessions. Each session lasted 40 minutes and was held once a week. The procedures employed were as follows: (1) Before the experiment, participants were instructed to use the systems to ensure they know how to use. (2) During these sessions, the number of participants’ attempted tasks were also recorded in the system logs as a supporting evidence for goal-pursuing questionnaire. This is because the number of attempted tasks could be regarded as an indicator of students’ goal-pursuing behavior. (3) At the end of each system version, the goal-pursuing questionnaire was administrated to collect students’ perceptions.

Results

Students’ perception of goal-pursuing

The questionnaire was administered to students subsequent to their participation in the experiment. The results of the questionnaire are presented in Table 2. The means of the via-version in three categories (e.g., enjoyment, goal orientation, and goal intensity) were higher than that of the without-version. To further validate the differences, the differences between the two system versions were found to be statistically significant by means of the t-test (t=2.39, p<.05; t=2.28, p<.05; t=2.21, p<.05 in the categories of enjoyment, goal orientation, goal intensity, respectively).
These differences indicated that participations in mathematics learning tasks using the via-version resulted in increased enjoyment, clearer goals, and stronger goal intensities.

Table 2. Summary table of t-test for questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Via-version</th>
<th>Without-version</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>3.56</td>
<td>1.023</td>
<td>3.07</td>
</tr>
<tr>
<td>Goal orientation</td>
<td>3.56</td>
<td>0.955</td>
<td>3.10</td>
</tr>
<tr>
<td>Goal intensity</td>
<td>3.41</td>
<td>1.008</td>
<td>2.95</td>
</tr>
</tbody>
</table>

* <.05

To have a comprehensive understanding of the two within-groups, the experiences of the participants in each group were also analyzed in terms of three aspects of the learning tasks. As illustrated in Table 3, measures of the three categories were significantly higher for Class A, which experienced the via-version prior to the without-version (t=2.35, p<.05; t=2.27, p<.05; t=2.32, p<.05 in enjoyment, goal orientation, and goal intensity, respectively). These results revealed that participants who engaged in learning tasks had better experiences in the via-version in terms of the three aspects. Moreover, as shown in Table 4, measures of the three categories were lower for Class B, which experienced the without-version prior to the via-version. When the differences were further examined by t-test, two categories were found to be significantly different (t=3.23, p<.05 in enjoyment; t=3.08, p<.05 for goal orientation), while no statistically significant difference was observed for the category of goal intensity (t=1.91, p>.05). This implied that participants who used the without-version later had relatively negative perceptions in terms of two aspects.

The results of Class A and B were combined, controlling for order effects and background influences. These results indicated that game quests have a positive impact on the perceptions of participants, especially in the aspects of enjoyment, goal orientation, and goal intensity. In other words, students reported more joyful learning experiences in the via-version, learned more information in completing the quest, and had stronger intentions to accomplish quests. A possible explanation for this is that quests promoted interaction with NPCs and involved more expectations and satisfactions, which, in turn, affected students' perception of quests towards goal-pursuing.

Table 3. Summary table of t-test for Class A

<table>
<thead>
<tr>
<th></th>
<th>Pre-test (without version)</th>
<th>Post-test (via-version)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>3.51</td>
<td>0.74</td>
<td>3.94</td>
</tr>
<tr>
<td>Goal orientation</td>
<td>3.54</td>
<td>0.94</td>
<td>3.89</td>
</tr>
<tr>
<td>Goal intensity</td>
<td>3.28</td>
<td>1.00</td>
<td>3.76</td>
</tr>
</tbody>
</table>

* <.05

Table 4. Summary table of t-test for Class B

<table>
<thead>
<tr>
<th></th>
<th>Pre-test (via-version)</th>
<th>Post-test (without-version)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>3.34</td>
<td>1.26</td>
<td>2.79</td>
</tr>
<tr>
<td>Goal orientation</td>
<td>3.39</td>
<td>1.21</td>
<td>2.89</td>
</tr>
<tr>
<td>Goal intensity</td>
<td>3.15</td>
<td>1.29</td>
<td>2.67</td>
</tr>
</tbody>
</table>

* <.05

Students’ attempted behavior

Table 5 illustrates attempted tasks with and without quests. The total number of attempted tasks in the via-version (=750) was significantly higher than that in the without-version (=572), t=4.13, p<.01. On average, each participant in the via-version attempted 17.04 (=750/44) quests; whereas each participant in the without-version attempted 13 (=572/44) quests. These results indicated that students were more motivated to acquire quests in the via-version. These results seemed to be consistent with feedbacks given by participants on the questionnaire.
Table 5. Summary table of t-test for attempted tasks

<table>
<thead>
<tr>
<th></th>
<th>Via-version</th>
<th>Without-version</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attempted tasks</td>
<td>750</td>
<td>572</td>
<td>t=4.13*</td>
</tr>
<tr>
<td>Total Mean</td>
<td>17.04</td>
<td>13.00</td>
<td></td>
</tr>
</tbody>
</table>

* < .05

Discussion

**Quests’ influence on students’ perception of task conduction**

The results showed that quests had positive impact on elementary school students, a claim that was supported by two types of evidence. As described above, task assignment was controlled such that tasks were either assigned directly or were assigned in the context of quests. The responses given by students to the questionnaires indicated that game quests are perceived more positively when they were introduced in the context of quests.

Furthermore, as more tasks were attempted in conditions with quests, this indicated that when students are more willing to undertake tasks in the context of quests. Although from a technical perspective, the delivery of learning tasks is ultimately the same either with or without quests, the perceptions and willingness of students to participate in learning activities are quite different. The findings seemed to indicate that quests produced more active participation and promoted more enjoyable experiences. Therefore, these results suggested that a student-centered perspective should be taken by designers of learning systems. Accordingly, the inclusion of quests would result in increased motivation to accomplish the tasks due to students’ perception of quest goals.

Moreover, game quests further involve another significant and interesting issue about the transfer between game goal and learning goal. More specifically, when learning tasks are embedded into the game quests, the learning goal seems to become a subordinate part in the pursuance of the game goal. Taking the two system versions as an example, in the without-version, students were assigned directly by the learning tasks. Their primary goal seems to complete the learning tasks; however, in the via-version, students received learning tasks via game quests, and their primary goal might be shifted to complete the game quests, in which the learning tasks become a side-effect for the primary goal. It is still unclear the influences of goal transfer on students’ learning, and more research is required in the future.

**Coupling mechanism: intrinsic or extrinsic approaches**

Another issue that merits discussion is the coupling mechanism used in the three-tier framework. In this study, we adopt quests as the coupling mechanism, which actually could be determined by internal or external approaches. An external approach means that quests are designed in accordance with a domain-independent approach and are more strongly correlated with extrinsic motivation. Intrinsic approaches, on the other hand, closely bind the quests with the specific lesson that is being learned, which may be considered as intrinsic motivation.

Regarding the nature of intrinsic and extrinsic approaches, however, Deci and Ryan (1985) indicated that educators “cannot always rely on intrinsic motivation to foster learning…because many of the tasks that educators want their students to perform are not inherently interesting or enjoyable.” In other words, an extrinsic approach might be also beneficial to students’ learning, since students who are extrinsically motivated to value and emphasize their tasks could become self-regulated and motivated towards actions. It is possible for students to cultivate their intrinsic motivation for subject matter, if they latter find that the learning process is an interesting and rewarding experience that is driven from extrinsic motivation to undertake the tasks at the initial stages.

In addition, a promising advantage that extrinsic approach could bring is the effect of the side-effect learning, students’ primary goal is naturally shifted as a game goal in which learning happened with the progress of the goal pursuing. Due to the flexibility of the extrinsic approach, learning contents in other subjects (e.g., language learning, science learning, or social studies) have more opportunities to be further integrated. For example, the knowledge of science learning or social studies could be integrated as a quest, which provides students with background knowledge or related information about science learning in the way of story-telling. In addition, understanding the content of
quests further involves students’ reading in language learning. It is feasible to integrate language learning pedagogies to promote students’ reading comprehension during the process of quest-taking, since narrative is a powerful method of information delivery, both for human culture transmission and for digital game design (Buckingham & Burn, 2007; Dickey, 2006).

**Conclusion**

This study shared an experience in developing a My-Pet-My-Quest game environment according to a three-tier framework. To answer the research question—what are the influences of game quests on students’ math learning in terms of perception of enjoyment and goal-pursuing, this study conducted an experiment focusing on the examination of the quest mechanism. Quests were found to exert significant influences on students’ perceptions, including enjoyment, goal orientation, and goal intensity. In addition, the findings revealed that quests encouraged active participation and increased the intensity of attempted learning behaviors.

However, due to the limitations of this experiment, some further investigations are required. Firstly, although this study showed positive influence of game quests on students, it was merely a short-term study. The long-term effects of quests are still unclear. In addition, because the questionnaire used in the experiment was developed by authors, its reliability and validity should be further examined in the future work. Finally, considering the ages of participants, other instruments (e.g., direct observation, interview or questionnaire to teachers) might be more precise to examine the effects of game quests. Other instruments could be taken into account in the future.

**Acknowledgements**

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financial support NSC-99-2631-S-008-001 and NSC-99-2511-S-008-002.

**References**


The Effect of Principals’ Technological Leadership on Teachers’ Technological Literacy and Teaching Effectiveness in Taiwanese Elementary Schools

I-Hua Chang
Department of Education, National Chengchi University, No. 64, Sec. 2, ZhiNan Rd., Taipei City 11605, Taiwan // ihchang@nccu.edu.tw

(Submitted January 17, 2011; Accepted March 8, 2011)

ABSTRACT

The purpose of this study is to investigate the relationships among principals’ technological leadership, teachers’ technological literacy, and teaching effectiveness. The survey target population consists of 1,000 teachers randomly selected from Taiwanese elementary schools. The survey asked teachers to measure the effectiveness of principals’ technological leadership, teachers’ technological literacy, and teaching effectiveness. Additionally, the data were analyzed using structural equation modeling to test how well the model captured the relationships between these quantities. The results of the common fit indices clearly indicate that the model fits the data quite well. The findings show that principals’ technological leadership improves teachers’ technological literacy and directly encourages teachers to integrate technology into their teaching. Furthermore, teachers’ technological literacy directly affects their effectiveness. Principals’ technological leadership also makes teachers more effective. Principals’ technological leadership, as mediated by teachers’ technological literacy, can affect teaching effectiveness. The results suggest that principals, as technology leaders, must develop and implement a vision and technology plan for their schools.

Keywords

Elementary schools, Teaching effectiveness, Technological leadership, Technological literacy

Introduction

The role of the principal has evolved from being primarily that of a building manager (Sharp & Walter, 1994) to that of an instructional and curricular leader (Checkley, 2000; Cheng, 2004; Glatthorn, 2000; Huang, 2004; Wu, 2004) and, more recently, to that of a technological leader (Anderson & Dexter, 2005; Bailey & Lumley, 1994; Ford, 2000; Inkster, 1998; Kadela, 2002; Matthews, 2002; McLeod, 2008; Scott, 2005; Seay, 2004; Stegall, 1998). Technological leadership is emerging within the increasingly diversified educational leadership world. Schools striving to excel in the information age need leaders that are well versed in the potential and in the pitfalls of information and communication technology. Many researchers (i.e., Anderson & Dexter; Byrom & Bingham; Gibson; Martin, Gersick, Nudell, & Culp) and educational organizations (i.e., the National School Boards Foundation; the United States Department of Education) note that strong leadership is essential to successful technology-based school reform (cited in Hughes, McLeod, Dikkers, Brahier, & Whiteside, 2005). Moreover, principals’ technological leadership strongly correlates with teachers’ integration of educational technology into their curriculums (Rogers, 2000). Technological leadership is vital for effective use of technology (Anderson & Dexter, 2005), and therefore, efforts to change and prepare schools and students for the information age demand effective technological leadership from principals (Ross & Bailey, 1996). As described above, in this era of digital technology, technological leadership is gaining importance. Thus, in pursuit of the ultimate goal of improving students’ abilities, principals aiming to facilitate school reform should have technological leadership abilities.

Technological leadership differs from traditional leadership theory in that it does not focus on the characteristics or actions of leaders but instead emphasizes that leaders should develop, guide, manage, and apply technology to different organizational operations so as to improve operational performance. Technological leadership is thus a type of functionally oriented leadership practice (Chin, 2010). The advancement and application of technology has infused new energy into educational reform. Examination of new educational reform plans in many countries shows that technological leadership has already become an important strategy for improving academic quality and student achievement. Taking the U.S. as an example, the education departments of many states have established technological leadership academies that use instructional programs to advance the technological professionalism of educators. The purpose of this is to enhance the instructional efficiency of teachers and the learning effectiveness of students (Chang & Tseng, 2005; ECS, 2001). The National Educational Technology Standards for Administrators...
(NETS-A) developed by the International Society for Technology in Education (ISTE) serves as a guide for the implementation of technological leadership in each state. Since 2004, Hong Kong has integrated the application of information technology for teaching and learning to leadership training courses (Chang, 2010). Likewise, the U.K. emphasizes the understanding and recognition of technology by school administrators, and encourages teachers to share their views on the use of information technology through cooperative methods (Robinson, 1994). Flanagan and Jacobsen (2003) suggested that inappropriate integration of technology results in negative by-products. This is because the importance of technological leadership lies not only in the use of technology but also in the development and change of school culture. Anderson and Dexter (2005) also indicated that the technological leadership of the school principal has a key influence on the effectiveness of technology utilization by teachers in educational instruction. Combining the above perspectives, this study used the viewpoint of technological leadership to examine problems with domestic education practices. Practically speaking, educators must realize that one-sided technology use cannot effectively enhance learning quality. Rather, only through the practical leadership in the school can an appropriate environment be built to change the mindsets of staff and create new instructional models. Internationally, the concept of technological leadership has received considerable emphasis, as indicated by studies (i.e., Anderson & Dexter, 2005; Bridges, 2003; Yee, 2000) and much research effort has been invested in investigating the practice and effectiveness of technological leadership.

In terms of policy, in 1980 the government of Taiwan began to place emphasis on investment in the application of information technology in education. “The Nine-Year Integrated Curriculum Policy” also emphasizes the development of basic technological skills. Ability to use information technology has become one of the new objectives of the school curriculum (Ministry of Education, 2003). “The 2008–2011 White Paper on Information Education in Junior High and Elementary Schools” from the Ministry of Education proposes eight main objectives and 20 strategies for the promotion of these objectives. One strategy is to encourage school principals and teachers to apply information technology in their teaching. Education departments in each city and county have made enhancement of information-technology literacy and usage an important focus in the professional growth of school principals. Taking “The Technology Leadership Training Workshop for Elementary School and Junior High School Principals” held by Department of Education, Taipei City Government in Taiwan in 2011 for example, such an activity is to strengthen school principals’ technological leadership, establish future vision, and improve school members’ technological competence through technological leadership, in order to effectively guide the establishment and management of technology facilities, encourage and support teachers to put technology into teaching practice, and cultivate the legal and ethical literacy for technology application. In addition, elementary school and junior high school principals can implement the promotion of information education policies and integrate the planning of school information technology with development, operation, and management-related issues. It is hoped that the training workshop can be used to provide the directions for understanding school information environment and integrating information technology with teaching and teachers’ professional development. Therefore, we can see that the concept of technological leadership has already taken root and has begun to grow in the field of education in Taiwan. The development of methods to effectively put technological leadership into practice to improve teaching and learning in schools is a critical issue to be highlighted by the principal.

**The purpose of the study**

The purpose of this study is to investigate the relationships among technological leadership of principals and the technological literacy and teaching effectiveness of elementary school teachers. The study also explores the notion that principals’ technological leadership, as mediated by teachers’ technological literacy, can affect teaching effectiveness.

**Theoretical framework**

**The dimensions of technological leadership**

Research on technological leadership began in the U.S. in the 1990s and has since gained significance. To be an experienced and capable technological leader, the principal must be trained in the following five areas:
(1) Vision, planning and management: As the most important foundation of technological leadership, a technological leader must develop a vision of how school reform will be affected by technology (Cory, 1990). The development of this vision requires that the principal understand the direction and trends of technology development, as this understanding will strongly influence the principal’s effectiveness. The principal must maintain a very clear technological vision, as well as understand the potential uses of technology in the classroom (Bailey, 1997; Bridges, 2003; Cory, 1990; Inkster, 1998; Jewell, 1998; Ray, 1992).

(2) Staff development and training: Ford (2000) showed that planning and establishing resources for staff development are the most important responsibilities of a technological leader. In staff development, the leader must prepare the newest models and material (Anderson & Dexter, 2000; Aten, 1996; Bailey, 1997; Bailey & Lumley, 1994; Ford, 2000; Inkster, 1998).

(3) Technological and infrastructure support: When instructors and staff require assistance, technological leaders must supply skilled support to preserve equal access to technological resources and appropriate technology-use environments. This must be a technological leadership skill of the principal (Anderson & Dexter, 2000; Aten, 1996; Bailey, 1997; Ford, 2000; Inkster, 1998; Kearsley & Lynch, 1994; Kline, 1993).

(4) Evaluation and research: Effective principals must administer procedures for measuring the growth of each individual teacher by rating instructors. They also must set technological targets and introduce professional development plans. An effective principal will gauge instructors’ performance by the results of studies on technological effectiveness. Simultaneously, principals should study students’ grades and encourage instructors to implement technology to improve academic performance (Aten, 1996; Cory, 1990; Ford, 2000; Inkster, 1998; ISTE, 1998, 2001; Kline, 1993; Moursund, 1992).

(5) Interpersonal and communication skills: Interpersonal skills are important in technological leadership, and these skills actually override technological skills. When new technology is implemented in a school, the leader must be able to provide support; thus, proper communication is the first skill that academic technology leaders must have. A complementary relationship is required between a principal’s communication skills and his or her closely correlated individual technological leadership skills (Aten, 1996; Bailey, 1997; Jewell, 1998; Moursund, 1992).

The dimensions of technological leadership have been drawn from the empirical literature on principals’ leadership in general and their effectiveness as technology leaders specifically. The aforementioned five dimensions have been examined and verified by early research (e.g., Chang, Chin, & Hsu, 2008; Chang, Hsiao, & Hsu, 2007; Chang & Hsu, 2009; Chang & Wu, 2008; Chin & Chang, 2006), and have been applied to relevant studies (i.e., Chang, 2006; Chang, 2009; Chen, 2006; Fu, 2009; Hsu, 2010; Lin, 2009; Lo, 2009; Sun, 2007; Tsai, 2008; Tsai, 2009; Tsai, 2010; Wu, 2006; Wu, 2008; Wu, 2009; Yen, 2010). The aforementioned five dimensions were chosen because they are the principals’ core tasks in dealing with teaching, learning, and administrative operations that involve technology in their schools.

**Technological leadership influences technological literacy**

Recently, regional departments of education have continued to provide instructors and administrators with information technology training or technological literacy training. Such training aims to increase instructors’ ability to fuse technology with academics and thereby increase students’ academic performance. In Chang and Hsu’s (2009) study, they explored the current state of elementary school principals’ technological leadership and teachers’ information-technology literacy in six urban areas of Taiwan. The results showed that principals’ technological leadership has a significant and positive influence on teachers’ information technology literacy (.42). Wu’s (2009) study explored the relationship between principals’ technological leadership and teachers’ technological literacy, as well as the differences in these factors in terms of background variables of individual teachers and in terms of school environmental variables. Next, the study analyzed the predictive power of principals’ technological leadership on teachers’ technological literacy. The study found a significant positive correlation overall. Principals’ technological leadership has predictive power for various aspects of teachers’ technological literacy.

The professional information technology knowledge, development, and training of administrators are important leadership factors. Important responsibilities in technological leadership lie in planning instructor training and
obtaining financial resources (Ford, 2000). In other words, classroom instruction is strongly correlated with planning and training designs stemming from effective technological leadership (ISTE, 1998). Rogers’ (2000) study showed that the Fort Wayne Community School instructors’ individual self-ratings for technology use in the classroom correlated with the knowledge and support for technology integration from the principal’s technological leadership. Chang’s (2004) study indicated that the principal’s information technological literacy was correlated strongly with instructor implementation of information technology. Thus, these researchers believe that the technological leadership of principals directly influences teachers’ technological literacy as well as their integration of information technology into the curriculum.

**Technological literacy influences teaching effectiveness**

The concept of teacher effectiveness comes from Bandura’s concept of self-efficacy, which focuses on an individual’s assessment of his or her own organizational and planning abilities before taking action to achieve a certain goal. Bandura believes that self-efficacy is a theory about changing actions; it is a medium for action recognition and a concept that includes elements such as outcome expectancy and efficacy expectancy. Using this theory in conjunction with the concept of teacher effectiveness, the outcome expectancy is the degree to which the teacher believes that a situation is controllable. The efficacy expectancy is the teacher’s assessment of his or her own ability to guide students to change positively. The phrase “teaching effectiveness” has several different definitions. Looking at its substance, we can analyze and discuss its main meaning based on the composition and level of the teacher’s reaction to students’ performance, students’ learning, and the teaching process. Teacher effectiveness is the degree to which teachers believe that they have the power to influence a student’s performance (Ashton, 1984). Teacher effectiveness is also the degree to which teachers believe that they themselves can influence a student’s learning (Gibson & Denbo, 1984); it is the belief that with the anticipation of special circumstances, teachers can help students learn (Ashton & Webb, 1986). Overall, teaching effectiveness can be seen as a teacher’s belief in his or her own teaching ability including teaching-material preparation, applications of teaching skills, classroom management, teaching assessment, and self-efficacy and belief.

Chang and Wu (2008) explored the current state of principals’ technological leadership and teachers’ teaching effectiveness in elementary schools in seven counties and cities in Taiwan. The results showed that principals’ technological leadership has a significant and positive effect on teachers’ teaching effectiveness (.77). Yen’s (2010) study explored the current state of principals’ technological leadership and teachers’ teaching effectiveness. The study found a positive correlation between the principals’ technological leadership as perceived by the teachers and the teachers’ teaching effectiveness; there was a significant predictive power of principals’ technological leadership for teachers’ teaching effectiveness and the factor of “staff development and training” was the most predictive. Fu’s (2009) study explored the current state and relationship between principals’ technological leadership and teachers’ teaching effectiveness and found a significant positive correlation. The factors “vision, planning, and management” and “evaluation and research” were the most predictive for overall teachers’ teaching effectiveness. Lo (2009) discussed the current state, differences, and correlations of principals’ technological leadership and teachers’ teaching effectiveness and also found a significant positive correlation; principals’ technological leadership can effectively predict teachers’ teaching effectiveness.

Eaton-Kawecki (2003) pointed out that the use of technology in school by both students and teachers can be used to accurately predict academic achievement. Clearly, a student’s academic achievements are often used to evaluate teaching effectiveness and are influenced by the use of technology in school. In other words, a student’s use of technology represents the teacher’s integration of technology into teaching and curricula and also affects the teacher’s effectiveness (for example, with regard to results of a statewide achievement performance assessment). Reed’s (2003) study showed that students’ academic achievements are noticeably influenced by the teacher’s use of technology. A teacher’s technological literacy directly affects whether students can incorporate technology into the curriculum to improve students’ academic achievements. Therefore, I propose that teachers’ technological literacy can improve teaching effectiveness. Actual evidence from studies (Jean, 2003; Pai, 2004) supports the perspective that teachers’ technological literacy affects teaching effectiveness. Is there a significant correlation between teachers’ technological literacy and teaching effectiveness? Does a principal’s technological leadership affect teaching effectiveness? In this paper, these questions will be discussed and explained.
Based on an examination of the literature on principals’ technological leadership and teachers’ technological literacy and teaching effectiveness, I hypothesize the model (see Figure 1) that principals’ technological leadership directly influences teachers’ technological literacy and improves teaching effectiveness. Furthermore, teachers’ technological literacy directly influences their teaching effectiveness. More important, the notion that principals’ technological leadership, as mediated by teachers’ technological literacy, can affect teaching effectiveness will be tested.

![Figure 1. Hypothesized conceptual model](image)

**Research method**

**Data collection**

The target population consisted of 1,000 teachers randomly selected from 100 elementary schools within the following six metropolitan cities in Taiwan: Keelung City, Taipei City, Hsinchu City, Taichung City, Tainan City, and Kaohsiung City. These cities are located in the northern, central, and southern parts of the island respectively, and this sample is indicative of the broader teaching population of Taiwan. The survey was randomly sent to the Director of Academic Affairs Division of each elementary school. Then, the Director randomly selected the teachers based on the cover letter sent to the school. The survey asked teachers to evaluate the principal’s role in leading and facilitating technology use, teachers’ technological literacy and teaching effectiveness in their schools. The respondents in this study were teachers (i.e., tenured teachers, mentor teachers, teachers serving as administrators and directors) reporting to their principals. To encourage the return rate, follow up thank you postcards and personal contacts were made until a satisfactory percentage of participants completed the questionnaire. Of the 1,000 questionnaires distributed, 60.5% were returned. Table 1 displays the demographic characteristics of the respondents. Demographic variables provide a descriptive profile of those individuals who responded to the survey.

*Table 1. Number and percent of respondents by demographics * N=605*

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>210</td>
<td>34.7%</td>
</tr>
<tr>
<td>Female</td>
<td>394</td>
<td>65.1%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 (or below) years old</td>
<td>118</td>
<td>19.5%</td>
</tr>
<tr>
<td>31-40 years old</td>
<td>268</td>
<td>44.3%</td>
</tr>
<tr>
<td>41-50 years old</td>
<td>186</td>
<td>30.7%</td>
</tr>
<tr>
<td>51 (or above) years old</td>
<td>33</td>
<td>5.5%</td>
</tr>
<tr>
<td>School Size (N of class)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 (or below) classes</td>
<td>104</td>
<td>17.2%</td>
</tr>
<tr>
<td>18-35 classes</td>
<td>172</td>
<td>28.4%</td>
</tr>
<tr>
<td>36-53 classes</td>
<td>178</td>
<td>29.4%</td>
</tr>
<tr>
<td>54 (or above) classes</td>
<td>150</td>
<td>24.8%</td>
</tr>
<tr>
<td>Teaching Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 (or below) years</td>
<td>120</td>
<td>19.8%</td>
</tr>
<tr>
<td>6-15 years</td>
<td>235</td>
<td>38.8%</td>
</tr>
<tr>
<td>16-30 years</td>
<td>224</td>
<td>37.0%</td>
</tr>
<tr>
<td>31 (or above) years</td>
<td>26</td>
<td>4.3%</td>
</tr>
<tr>
<td>Educational Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>BA or BS</td>
<td>97</td>
<td>16.0%</td>
</tr>
<tr>
<td>BA or BS (from teachers college or normal university)</td>
<td>338</td>
<td>55.9%</td>
</tr>
<tr>
<td>M.A. or M.Ed.</td>
<td>165</td>
<td>27.3%</td>
</tr>
<tr>
<td>Ph.D. or Ed.D.</td>
<td>1</td>
<td>0.2%</td>
</tr>
<tr>
<td>Position</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mentor Teacher</td>
<td>256</td>
<td>42.3%</td>
</tr>
<tr>
<td>Specialist Teacher</td>
<td>69</td>
<td>11.4%</td>
</tr>
<tr>
<td>Administrative Head</td>
<td>190</td>
<td>31.4%</td>
</tr>
<tr>
<td>Director</td>
<td>88</td>
<td>14.5%</td>
</tr>
</tbody>
</table>

* Numbers and percentages in the categories may not total 100 due to some missing data.

**Instrumentation and variables**

*The Principals' Technological leadership Instrument* was conceptualized as five interrelated dimensions: (1) vision, planning and management (e.g., clearly articulates a shared vision for technology use in the school); (2) staff development and training (e.g., encourages technology in-service training); (3) technological and infrastructure support (e.g., ensures appropriate technology facilities); (4) evaluation and research (e.g., considers effective technology use as one performance assessment component of instructional staff); and (5) interpersonal and communication skills (e.g., demonstrates an understanding of technology needs and concerns of faculty). *The Teachers' Technological Literacy Instrument* comprised four dimensions: (1) hardware and software operation (e.g., understands the functions of basic components of computer hardware); (2) law and ethics (e.g., understands the meaning of ownership and publication, and show respect to intellectual property rights); (3) technological integration (e.g., understands the benefits and scope for applying computer to teaching); and (4) management and assessment (e.g., possesses the ability in using teaching software and evaluating online teaching materials). *The Teachers' Teaching Effectiveness Instrument* included five dimensions: (1) teaching-material preparation (e.g., be comprehensively familiar with the content of the unit to be taught before teaching); (2) applications of teaching skills (e.g., continuously attracts students’ attention by changing activities during class); (3) classroom management (e.g., continually compliments and encourages students for their progress); (4) teaching assessment (e.g., effectively implements various evaluation methods according to teaching needs); and (5) self-efficacy and belief (e.g., has a positive influence on students’ learning achievements). These three instruments comprised 101 Likert-type items on 5-point scales. A response of “1” indicated that the teacher strongly disagreed with the statement on the scale, and a response of “5” indicated that the teacher strongly agreed with the statement on the scale. A pilot study was conducted with approximately 204 teachers and school administrators to improve the instruments’ reliability and validity. The Cronbach’s alpha coefficients (shown in parentheses and see also table 2) were calculated for each scale: vision, planning and management (.96); staff development and training (.90); technological and infrastructure support (.92); evaluation and research (.96); interpersonal and communication skills (.93); hardware and software operation (.94); law and ethics (.90); technological integration (.91); management and assessment (.90); teaching-material preparation (.92); applications of teaching skills (.88); classroom management (.93); teaching assessment (.89); and self-efficacy and belief (.95).

In addition to reliability analysis, construct validity tests of the aforementioned three instruments were performed. Five factors were extracted using Varimax rotation from *The Principals' Technology Leadership Instrument*. The variance factors were as follows: 21.148% (vision, planning and management), 15.858% (evaluation and research), 15.430% (technological and infrastructure support), 14.573% (interpersonal and communication skills), and 11.958% (staff development and training). The total explained variance of these five factors was 78.966%. Four factors were extracted using Varimax rotation from *The Teachers' Technology Literacy Instrument*. The variance factors were as follows: 24.474% (hardware and software operation), 22.055% (technological integration), 13.304% (law and ethics), and 10.375% (management and assessment). The total explained variance of these four factors was 70.208%. Five factors were extracted using Varimax rotation from *The Teachers' Teaching Effectiveness Instrument*. The variance factors were as follows: 22.082% (applications of teaching skills), 19.060% (self-efficacy and belief), 12.726% (teaching-material preparation), 11.314% (teaching assessment), and 10.029% (classroom management). The total explained variance of these five factors was 75.211% (see table 2).
Table 2. Analysis of reliability and validity

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Cronbach’s alpha</th>
<th>Variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision, planning and management</td>
<td>.96</td>
<td>21.148%</td>
</tr>
<tr>
<td>Staff development and training</td>
<td>.90</td>
<td>11.958%</td>
</tr>
<tr>
<td>Technological and infrastructure support</td>
<td>.92</td>
<td>15.430%</td>
</tr>
<tr>
<td>Evaluation and research</td>
<td>.96</td>
<td>15.858%</td>
</tr>
<tr>
<td>Interpersonal and communication skills</td>
<td>.93</td>
<td>14.573%</td>
</tr>
<tr>
<td>Hardware and software operation</td>
<td>.94</td>
<td>24.474%</td>
</tr>
<tr>
<td>Law and ethics</td>
<td>.90</td>
<td>13.304%</td>
</tr>
<tr>
<td>Technological integration</td>
<td>.91</td>
<td>22.055%</td>
</tr>
<tr>
<td>Management and assessment</td>
<td>.90</td>
<td>10.375%</td>
</tr>
<tr>
<td>Teaching-material preparation</td>
<td>.92</td>
<td>12.726%</td>
</tr>
<tr>
<td>Applications of teaching skills</td>
<td>.88</td>
<td>22.082%</td>
</tr>
<tr>
<td>Classroom management</td>
<td>.93</td>
<td>10.029%</td>
</tr>
<tr>
<td>Teaching assessment</td>
<td>.89</td>
<td>11.314%</td>
</tr>
<tr>
<td>Self-efficacy and belief</td>
<td>.95</td>
<td>19.060%</td>
</tr>
</tbody>
</table>

Data analysis

This research empirically investigated the relationships among principals’ technological leadership, teachers’ technological literacy, and teaching effectiveness and tested the structural equation model (SEM). SPSS was used to calculate scale reliabilities (the Cronbach’s alpha) and to perform factor analyses. The final SEM model was done with LISREL.

Results

SEM lets researchers simultaneously define and measure multidimensional constructs (e.g., principals’ technological leadership). The fit of the model to the data was assessed with SEM fit indices. The Chi-Square value was 271.97 with 74 degrees of freedom and significance \( p = .00 \). The NFI, NNFI, PNFI, CFI, SRMR, GFI, AGFI and PGFI values were .98, .99, .80, .99, .043, .93, .90 and .66, respectively. The Critical N was 204.42. The results of these common fit indices (see table 3) clearly indicated that the model fit these observed data quite well.

Table 3. Assessment of overall model fit

<table>
<thead>
<tr>
<th>Fit Indices</th>
<th>Ideal fit index</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of Freedom=74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute fit indices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Square</td>
<td>Not significant</td>
<td>271.97 ( p = .00 )</td>
</tr>
<tr>
<td>GFI</td>
<td>&gt; .90</td>
<td>.93</td>
</tr>
<tr>
<td>AGFI</td>
<td>&gt; .90</td>
<td>.90</td>
</tr>
<tr>
<td>SRMR</td>
<td>≦ .05</td>
<td>.043</td>
</tr>
<tr>
<td>Comparative fit indices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNFI</td>
<td>&gt; .90</td>
<td>.99</td>
</tr>
<tr>
<td>CFI</td>
<td>&gt; .90</td>
<td>.99</td>
</tr>
<tr>
<td>NFI</td>
<td>&gt; .90</td>
<td>.98</td>
</tr>
<tr>
<td>Parsimonious fit indices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNFI</td>
<td>&gt; .50</td>
<td>.80</td>
</tr>
<tr>
<td>PGFI</td>
<td>&gt; .50</td>
<td>.66</td>
</tr>
<tr>
<td>Critical N</td>
<td>&gt; 200</td>
<td>204.42</td>
</tr>
</tbody>
</table>

After ensuring that the model fit the data, the standardized parameter estimates were considered in the model. As shown in Figure 2, the parameter estimates for the five constructs (i.e., vision, planning and management; staff development and training; technological and infrastructure support; research and evaluation; interpersonal and
communication skills) that comprised principals’ technological leadership were significant (.91, .82, .89, .90, and .87). The parameter estimates for the four constructs (i.e., hardware and software operation; law and ethics; technological integration; management and assessment) that comprised teachers’ technological literacy were significant (.83, .85, .88, and .92). The parameter estimates for the five constructs (i.e., teaching-material preparation; applications of teaching skills; classroom management; teaching assessment; and self-efficacy and belief) that comprised teaching effectiveness were significant (.83, .92, .90, .88, .90). As presented in Figure 2, principals’ technological leadership (TECHLEAD) had a significant positive effect on teachers’ technological literacy (LITERACY) (.58).

![Figure 2. Final structural equation model](image)

These findings confirmed previous research that principals’ technological leadership improves teachers’ technological literacy development and directly influenced teachers’ integration of technology into their teaching practices. Furthermore, teachers’ technological literacy (LITERACY) directly affected teaching effectiveness (EFFECTIVE) (.65). Principals’ technological leadership also improved teachers’ teaching effectiveness (.22). The findings also supported previous research mentioned earlier in this study. Table 4 displays the summary of parameter estimation for the mediated-effects model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>t</th>
<th>R²</th>
<th>Standardized solution</th>
<th>Parameter</th>
<th>t</th>
<th>R²</th>
<th>Standardized solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ₁</td>
<td>-</td>
<td>0.83</td>
<td>0.91</td>
<td>δ₁</td>
<td>12.05</td>
<td>-</td>
<td>0.18</td>
</tr>
<tr>
<td>λ₂</td>
<td>26.07</td>
<td>0.67</td>
<td>0.82</td>
<td>δ₂</td>
<td>14.34</td>
<td>-</td>
<td>0.33</td>
</tr>
<tr>
<td>λ₃</td>
<td>31.54</td>
<td>0.79</td>
<td>0.89</td>
<td>δ₃</td>
<td>12.66</td>
<td>-</td>
<td>0.20</td>
</tr>
<tr>
<td>λ₄</td>
<td>32.30</td>
<td>0.81</td>
<td>0.90</td>
<td>δ₄</td>
<td>12.30</td>
<td>-</td>
<td>0.19</td>
</tr>
<tr>
<td>λ₅</td>
<td>29.88</td>
<td>0.76</td>
<td>0.87</td>
<td>δ₅</td>
<td>13.31</td>
<td>-</td>
<td>0.24</td>
</tr>
<tr>
<td>λ₆</td>
<td>-</td>
<td>0.69</td>
<td>0.83</td>
<td>ε₁</td>
<td>13.41</td>
<td>-</td>
<td>0.31</td>
</tr>
<tr>
<td>λ₇</td>
<td>16.39</td>
<td>0.42</td>
<td>0.65</td>
<td>ε₂</td>
<td>15.25</td>
<td>-</td>
<td>0.57</td>
</tr>
<tr>
<td>λ₈</td>
<td>25.20</td>
<td>0.77</td>
<td>0.88</td>
<td>ε₃</td>
<td>11.65</td>
<td>-</td>
<td>0.22</td>
</tr>
<tr>
<td>λ₉</td>
<td>26.91</td>
<td>0.85</td>
<td>0.92</td>
<td>ε₄</td>
<td>8.97</td>
<td>-</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Discussion

Technological leadership’s effects on technological literacy and teaching effectiveness

As presented in Figure 2, principals’ technological leadership (TECHLEAD) improves teachers’ technological literacy (LITERACY) significantly. These findings confirm previous research that principals’ technological leadership improves teachers’ technological literacy development (Chang & Hsu, 2009; Wu, 2009) and directly influences teachers to integrate technology into their teaching (Chang, 2004; Eaton-Kawecki, 2003; Reed, 2003; Rogers, 2000). Furthermore, teachers’ technological literacy (LITERACY) directly improves their effectiveness (EFFECTIVE) as does principals’ technological leadership (TECHLEAD) directly influences their effectiveness (EFFECTIVE) as does principals’ technological leadership (Chang & Wu, 2008; Fu, 2009; Jean, 2003; Lo, 2009; Pai, 2004; Yen, 2010). The findings also support previous research mentioned earlier in this study. Based on the aforementioned results, schools must develop a long-term technology plan with a vision and goals for the school’s faculty and staff. In addition, development of technological skills is essential for the professional growth of faculty and staff. Regarding the provision of infrastructure, schools must provide full technological support and equal opportunities to obtain resources. Lastly, schools must be able to assess themselves; the leaders must be able to evaluate their own professional technological plans and give teachers performance evaluations. Principals’ interpersonal and communication skills subtly affect the success of their leadership. Principals who can listen to teachers’ advice and opinions while providing them with sufficient support will make their school’s operational development more productive. Only a principal who can understand current knowledge of technology can ensure that the school will continue to develop. In addition, when a principal is competent at technological leadership and integration and also creates a technological learning environment, teaching effectiveness will increase and in turn will improve student performance.

Principals’ technological leadership mediated by teachers’ technological literacy can affect teaching effectiveness

Based on the results of the literature review indicating that teachers’ technological literacy mediates the effects of principals’ technological leadership on teaching effectiveness, the mediated-effects model was proposed (see Figure 1). As shown in Figure 2, principals’ technological leadership (TECHLEAD) positively and significantly ($\gamma_{11} = .58$, $p < .05$; $\beta_{21} = .65$, $p < .05$) influences teachers’ teaching effectiveness (EFFECTIVE). Principals’ technological leadership (TECHLEAD) also positively and significantly ($\gamma_{11} = .58$, $p < .05$; $\beta_{21} = .65$, $p < .05$) influences teachers’ teaching effectiveness (EFFECTIVE) via teachers’ technological literacy (LITERACY). The indirect effects value is .38 ($\gamma_{11}^*\beta_{21}$), and the R-Squared value is 0.64. The result shows that through the mediated-effects of teachers’ technological literacy, the principals’ technological leadership can explain 64% of the variance in teaching effectiveness. The result of the mediated-effects model indicates that teachers’ technological literacy results in significant mediated effects of principals’ technological leadership on teaching effectiveness.

Conclusions

As this study indicates, principals as technological leaders must develop and implement vision and technology plans for their schools, encourage the technological development and training of teachers, provide sufficient technological infrastructure support, and develop an effective school-evaluation plan. Principals who can embrace their ever-
changing roles and become technological leaders are those who can effectively lead and prepare their schools for the decades to come. More important, teachers, who practice their craft in the technological age, should look diligently to develop their own technological literacy; they should change the traditional unidirectional method of teaching, and learn how to incorporate technology into their teaching in order to encourage their students’ academic achievement. It is an open question whether, in this age of information, principals (with their ever-evolving roles) can become competent technology leaders who nurture and increase the ability of teachers to integrate technology into their teaching. Such leaders not only make teachers more effective but also directly affect students’ academic achievement.

Implications for practice

A school’s internal administration is no longer a closed-circuit system but rather a dynamic educational ecology. Schools can no longer ignore the influence of technological development outside of the school. In particular, schools in the information age can no longer face the rapidly changing world with an attitude that hopes to preserve the past. Similarly, the role of the principal has also changed from solely a school administrator to the current, multi-faceted role of curricular and technological leader. Therefore, the most important task of a principal is to figure out how to become an appropriate technological leader. A principal can then guide teachers to improve their technological literacy (or technology-implementation abilities) and to improve teaching effectiveness and students’ academic achievements through educational reforms. This study acts as a reference for school leaders regarding administrative management. One important thing for principals to remember is that while they are carrying out technological leadership, they should not focus only on hardware. There is a need to emphasize not only the involvement of technology but also the integration of technology.

Implications for theory

The study uses structural equation modeling to test the relationships among principals’ technological leadership, teachers’ technological literacy, and teaching effectiveness. In addition to technological leadership positively influencing teaching effectiveness, the results show that principals’ technological leadership influences teaching effectiveness through the mediated-effects of teachers’ technological literacy. When interpreting the relationships among the three variables, the function of the “mediated-effects” in the model should be emphasized. That is, technological leadership influences teaching effectiveness through technological literacy.

Acknowledgements

This work was supported by grant from the National Science Council (NSC) of the Republic of China, Taiwan (Grant no. 94-2413-H-004-024). I would like to thank the NSC for financial support, and the teachers who are the most important participants in this study.

References


Hsu, P. (2010). A study of the relationship between elementary/junior high school administrators’ perceptions on technological leadership and minority students’ information literacy (Unpublished master’s thesis). Tamkang University, Taiwan.


Instructor Learning Styles as Indicators of Online Faculty Satisfaction

Ryan McLawhon¹ and Marc Cutright²

¹Texas A&M University, MS 1157 College Station, TX 77843, USA // ²University of North Texas, 1155 Union Circle #310829 Denton, TX 76203-5017, USA // ryan.mclawhon@tamu.edu // Marc.Cutright@unt.edu

(Submitted April 15, 2011; Revised June 23, 2011; Accepted July 8, 2011)

ABSTRACT
The purpose of this study was to determine the relationship between instructor learning style/preference and online faculty job satisfaction. Learning style was assessed using the Readiness for Education At a Distance Indicator (READI) now called Smarter Measure. Online faculty job satisfaction was assessed using the National Study of Postsecondary Faculty (NSOPF) job satisfaction questions. Analysis of variance was used to determine whether there was a difference in satisfaction based on individual instructor learning preference(s). The sample population (N=110) consisted of online faculty members in one large community college district. A significant difference in satisfaction with equipment and facilities was found based on instructor learning preference. Implications and recommendations for future practice and research are presented, including considerations for international application.

Keywords
Faculty satisfaction, Learning styles, Distance learning, Online faculty

Introduction

Prior to teaching online, many instructors predict future dissatisfaction with the online classroom. These predictions are usually based on perceptions of technical skills, personality type, unfamiliarity, etc. Further, many instructors have not or could never picture themselves enjoying or taking an online class as a student. Like students, they may say, “I just don’t learn that way.” However, these are merely perceptions of potential dissatisfaction.

All of these could indeed be indicators of online faculty satisfaction. However, this discussion has been based on perceived levels of future satisfaction in the online classroom. A more valid and worthwhile discussion would contain actual levels of satisfaction in the online environment following the actual teaching experience. Once the levels are known, comparing them to instructor learning styles would provide deep insight to understanding online faculty satisfaction.

Faculty retention is an issue with which many institutions are concerned. Certainly, job satisfaction contributes to a faculty member’s intent to leave (Rosser & Townsend, 2006). However, with deeper implications, faculty job satisfaction contributes to the quality of work performed (Katzell, Thompson, & Guzzo, 1992). The quality of work performed by the faculty at an institution affects numerous areas. Perhaps the most important stakeholders affected by high faculty performance are the students. The lifeline of any institution is its students; therefore, retaining those who enroll, thus reducing attrition, is one of the most important tasks in maintaining institutional effectiveness.

John Bean (2005) identified one of the nine themes of college student retention as academics. Within this discussion is Bean’s understanding that faculty members are the ones who deliver the institution’s product. He stated that, “A faculty member presents substantive material in a course in a way that promotes or does not promote students to be socialized to academic values and choose a particular major” (Bean, p. 225). For Bean, the quality of the interaction between the student and the faculty member directly contributes to his or her affinity towards the institution. Braxton and Hirschy (2005) also understand the relationship between quality faculty work and student persistence. Based on the work of Vincent Tinto, they suggested that faculty members who involve students in the learning process by engaging them in critical thinking about content contribute to persistence and retention. Engaging students in their academic experience is not only a model of quality teaching, but it is also of extreme importance in commuter campuses such as community colleges. Higher levels of academic involvement leads to higher levels of institutional commitment; institutional commitment in turn leads to persistence (Braxton, & Lien, 2000).

Research on online instructor satisfaction is extremely limited in the field of higher education. Additionally, little to no research has been completed on individual online instructor learning styles. Most of the research on online
instruction has studied the student learner; these studies usually list individual instructor personality, ability, and style as delimitations. To avoid these delimitations, other studies attempt to hold the instructor variable constant; however, this method results in an extremely small sample size, resulting in less significant results. Because this study focused on areas virtually unstudied and undiscovered, the results are extremely helpful. This study should open new avenues for online education research, shifting some of the focus away from the student to the online instructor’s readiness.

The findings of this study offer implications in a number of areas. First, the results provide academic departments insight on who to best choose for online course development and teaching. Next, the findings offer suggestions for online faculty training and professional development personnel. Also, it is possible that these findings could offer insight into the realm of student retention strategies (as they relate to quality instruction). Lastly, this will seek to address issues related to faculty retention and intent to leave.

**Related literature and theoretical framework**

Because of the subjective nature of a large portion of this study, a review of the theoretical literature related to the discussion is necessary. It is important to note that this literature review addresses issues related to satisfaction that are not necessarily specific to online instruction. Rather, this review is more concerned with faculty as a whole.

**Theory**

This discussion is not an attempt to address the question of what it means to be satisfied; rather, this discussion focuses solely on job satisfaction as it relates to faculty in higher education institutions. Ask ten different faculty members what it means to be satisfied with their job in higher education and there will be ten different responses. However, there are several points of consensus on the issue that are necessary to understanding the essence of this mostly subjective concept.

**Frameworks**

Multiple scholars have developed frameworks for understanding job satisfaction. These frameworks not only address what it means to be satisfied with one’s job but also how satisfaction is attained. This section will introduce a brief overview of the primary contributors to job satisfaction theory.

**Maslow’s hierarchy of needs**

For Maslow (1954), job satisfaction is attained when the job and its environment meet the needs of the individual. Needs are organized hierarchically with physiological, safety, belongingness, and love needs holding the lower levels. The higher levels are filled with needs associated with esteem and self-actualization. The lowest unmet level serves as the primary need; once that need is fulfilled, needs on the next level up surface as the predominant needs. Motivations shift to meet newly surfaced needs as other needs are met.

**Vroom’s expectancy theory**

Vroom (1964) built upon the work of Maslow, adding a level of human choice to the equation. Here, individuals make decisions about their work based on their perceived abilities to successfully perform the tasks and receive the reward. Decisions are made considering three variables: expectancy (perceived ability), instrumentality (connection between success and reward), and valence (value of the expected reward). When all three variables achieve a high level, motivations rise commensurately, and subsequently, so do performance choices.
Herzberg’s motivator-hygiene theory/Hagedorn

In her helpful section in the book she edited, *What Contributes to Job Satisfaction among Faculty and Staff*, Linda Serra Hagedorn (2000) offered a two-construct model for conceptualizing faculty job satisfaction based primarily on the work of Frederick Herzberg; the two constructs that affect job satisfaction were triggers and mediators. Hagedorn defined a *trigger* as a “significant life event that may be either related or unrelated to the job” (p. 6). Because these triggers could affect the entire self and view of life, they could affect job satisfaction subsequently. She defined a *mediator* as a “variable or situation that influences (moderates) the relationship between variables or situations producing an interaction effect” (Hagedorn, p. 6). Mediators included motivators and hygienes (demotivators), demographics, and environmental conditions, while triggers included change in life stage, change in family-related or personal circumstances, change in rank or tenure, transfer to a new institution, change in perceived justice, or change in mood or emotional state. Herzberg’s model is helpful for the purpose of categorizing contributors to faculty satisfaction; however, it hardly addresses the difficulty of quantifying the issue.

Satisfaction and the institution

In an effort to quantify faculty satisfaction, Elizabeth Pollicino (1996) presented a paper to the American Educational Research Association in which she sought to intertwine the complexity of satisfaction with professional values as they relate to institutional values. She, therefore, defined *satisfaction* as, “the extent to which faculty perceive that the institution provides a climate ensuring professional autonomy and activity commensurate with specialized expertise” (Pollicino, p. 4). Pollicino’s study rested upon the theoretical assumption that “to the extent that faculty perceive institutional support for their professional activities they will likely derive satisfaction from their work and manifest loyalty to their employing institution” (p.3). She, therefore, placed the onus of high satisfaction levels on the shoulders of the institution; she assumed it is the institution’s responsibility to foster a “climate conducive to faculty satisfaction” (Pollicino, p. 4). Faculty ownership of the institutional values is the only level at which individual faculty member responsibility lie in her model. In comparison, the Herzberg categorization model places the bulk of the faculty satisfaction responsibility on the individual and his or her life events. While this is not an either/or discussion, it is interesting to note the varying degrees of locus of control (Pollicino removed most of the locus of control from the individual). These two frameworks are not mutually exclusive; Herzberg did include a level of institutional culture within her “mediator” construct. The difference is that Herzberg more closely related satisfaction to personal and partially unpredictable variables, while Pollicino focused on institutional characteristics.

Motivation

As presented by Maslow (1954) and Vroom (1964), essential to the discussion of satisfaction is the concept of motivation. Motivators can be both intrinsic and extrinsic. For example, Pollicino (1996) placed the importance of motivation, as it relates to satisfaction, extrinsically on the institution. As presented by Vroom, the extent to which one is motivated to perform a certain task directly relates to the satisfaction gained from achieving that task. Jerry Berberet (2006), in his contribution to Bataille and Brown’s *Faculty Career Paths*, reported that faculty are motivated most by four factors: 1) the intellectual stimulation they receive from their field of study, 2) the desire to be a member of a meaningful academic community, 3) the goal to make teaching, research, and service contributions, and 4) the wish to be recognized for their contributions. (p. 114)

Berberet reported that the primary factor of the four was intellectual stimulation followed by institutional recognition.

Because satisfaction is so closely related to motivation, further study of motivational constructs and processes is necessary. Robert Blackburn and Janet Lawrence (1995) grouped motivation theories into two categories: noncognitive and cognitive. Further, in an effort to quantify levels of motivation, Blackburn and Lawrence focused on “situations in which there are performance outcomes that define levels of success” (p. 18). For Blackburn and Lawrence, noncognitive theories assumed that behavior can be predicted based on internal needs, personality, and rewards/incentives. Cognitive theories assume that behavior is a result of individual perception of capacity to
respond and the estimation of possible losses and gains. In summary, individual motivation is the balancing of self-efficacy, personality, and perceived rewards; here, Blackburn and Lawrence combined the work of Maslow (1954) and Vroom (1964).

**Importance of faculty satisfaction**

Up to this point, the importance of high levels of faculty satisfaction within an institution has only been implied. It is easy to assume that institutions want their faculty to be satisfied with their jobs. However, it is necessary to outline the relevance of job satisfaction and what it means for the institution.

**Intent to leave**

Most of the faculty satisfaction research points to correlations between satisfaction and intent to leave as the primary concern. Logically, institutions desire a high level of faculty retention, especially among those faculty members producing a desired level of quality work (addressed below). In their 2006 study, Rosser and Townsend drew on the work of Herzberg to “develop a model that simultaneously defines multidimensional constructs such as worklife, job satisfaction, and intent to leave” (p. 129). Rosser and Townsend used the results from the National Study of Postsecondary Faculty (NSOPF-1999) as their data source. Of special importance to this study is the construct of worklife; this construct contains three dimensions: technical support, administrative support, and professional development. The conclusion of the study was that worklife factors significantly affect satisfaction and satisfaction negatively affects intent to leave.

**Quality of work**

In addition to retention, faculty satisfaction affects quality of work; this concept draws upon the discussion concerning motivation presented above. Additionally quality of faculty work is of extreme importance regarding student persistence and retention. Highly satisfied faculty members experience higher levels of motivation to perform their duties. Katzell et al. (1992) conducted research on multiple departments within four organizations ($n=1,200$) and found, while job satisfaction had little or no effect on self-assessed job performance, it significantly influenced job performance as evaluated by superiors ($r=.17$). Katzell et al. also reported a strong correlation between job satisfaction and job involvement ($r=.37$); job involvement influenced level of effort, which in turn strongly correlated with self-assessed job performance ($r=.49$).

**Faculty satisfaction measures**

During the time of data collection, the NSOPF was the most widely used tool to measure faculty satisfaction in higher education. The survey however has not been administered nationwide since 2004. Other tools, such as the Faculty Survey of Student Engagement (FSSE), offer insight into understanding faculty satisfaction. However, the FSSE simply measures faculty perceptions of student involvement and motivation. Many institutions such as Hudson Valley Community College (2011) have designed internal measures of faculty satisfaction. Hudson Valley’s Online Faculty Satisfaction Survey targets online faculty specifically. Likewise, Bolliger, Doris and Wasilik (2009) developed the Online Faculty Satisfaction Survey (OFSS) to “develop and validate an instrument that can be used to measure perceived faculty satisfaction in the context of the online learning environment” (p. 103). While many other institutions and researchers have developed similar tools, no one instrument has been as widely distributed as the NSOPF.

**Learning styles/preferences**

Most scholarly research on learning styles focus heavily on the learner. However, little research has been conducted on how an individual faculty member’s learning style affects his or her teaching style. The general assumption is that teachers teach based on their personal learning preferences while attempting to take into consideration the known or
perceived learning style(s), of the students. Veronica and Lawrence (1997) conducted a study to assess the affect of secondary school teachers’ learning preferences on class management and teaching methodology in which they found that the majority of teachers in their sample were identified as Reflector learners; these learners tend to demonstrate watching and thinking actions in a learning environment. Veronica and Lawrence recorded that regarding class management, many were comfortable teaching such that they controlled both the information and the way their students were expected to learn. These characteristics fall within the category of the Reflector preference within the Honey and Mumford learning style model.

In the study by Fuller et al. (2000) mentioned above, the Transaction Ability Inventory was implemented to assess the relationship between teaching styles and probability to proceed with online instruction. They found that Abstract Random teaching styles were more conducive to intend to continue teaching online. These teachers incorporate flexibility, appreciate individuality, enjoy group work, and stress the importance of the affective domain. Concrete Sequential learners are less likely to continue teaching online; they use step-by-step processes, require order, use facts and details, incorporate guided practice and tactile kinesthetic pedagogical tools, and focus on the cognitive domain.

**Online readiness measures**

The READI Assessment is widely used in higher education to assess student readiness for the online environment. Some institutions have internally developed tools designed to assess similar constructs, such as the University System of Georgia’s Online Readiness Tool and the University of Illinois at Springville’s Technology-Enhanced Learning – Browser test. Some of these internal tools have been adopted by other institutions such as Kerr, Ryneron, and Kerr’s (2006) Test of Online Learning Success (TOOLS). It is important to note that most, if not all, of these tools were designed to measure student online readiness, not faculty readiness.

Further, several learning style/preference measures are available, and are usually embedded within the available online readiness tools. In addition to measures based on Howard Gardner’s work, tools such as Clifton and Anderson’s (2002) Strengthsquest and the Learning Style Inventory based on Kolb’s (1984) Experiential Learning Theory (ELT) are also helpful in understanding learning preference. However, these measures also focus on the student rather than the instructor.

**Current state of knowledge**

As presented above, the current state of knowledge on job satisfaction, both theoretical and research-based, is vast. Likewise, studies pertaining to higher education faculty satisfaction are readily available. Most of these studies incorporate the findings of national surveys such as the Department of Education’s Survey of Postsecondary Faculty or the Carnegie Foundation’s National Survey of Faculty. A primary area of research that is currently missing in higher education scholarship is that of online or distance learning faculty satisfaction. There is little to no research documenting satisfaction indicators as they relate to online instruction specifically. Surely, based on the presentation of literature and research above, higher education scholarship would benefit from such a study.

Likewise, there is quite a bit of research discussing multiple implications for online instructor learning preferences. However, most of these studies, with a few exceptions, fall short of addressing their relationship to faculty satisfaction, much less online faculty satisfaction. Given the current culture of online learning and instruction and the importance of understanding faculty satisfaction, relationships between the two must be sought and understood.

**Design**

This was a quantitative study concerned with analysis of variance. The results were used to describe relationships between online faculty learning styles and job satisfaction. Learning style was the independent variable, which was assessed using the READI Assessment. Job satisfaction was the dependent variable and was assessed using the NSOPF job satisfaction questions.
Research question

For the purpose of describing job satisfaction variance among online faculty members, this study was designed to address the following research question:

*Is there a statistically significant difference in online instructor job satisfaction based on individual instructor learning style?*

Design and data collection

This study was limited to online instructors in one large community college district (five campuses). During the time of the study there were approximately 39,000 enrollments at the college with close to 25% of students taking at least one online course. An online instructor was defined as an individual instructor who teaches at least one course in a given semester fully online. Some of these instructors also teach face-to-face during each semester as the institution does not employ full-time online faculty. Each online instructor had completed the same required training and course development program.

Assumptions

There were three primary assumptions of this study. First, every online instructor at the institution has completed the same level of online faculty training provided by the district. This training not only focused on their technical ability to build a course and maintain the course management system, but also on online pedagogical theory and strategies. Second, this study assumed, based on research presented below, that the READI Assessment, although designed to assess online student readiness, accurately assesses online instructor learning styles. Third, this study assumed that the National Study of Postsecondary Faculty job satisfaction questions accurately assess levels of faculty satisfaction.

Research respondent selection

All online instructors at the institution were sent an email informing them that their participation in the study was requested. The email contained a brief introduction and a Web link to access the research Web site. The research Web site was designed to explain the study in detail and provide information regarding access to the study, including the Informed Consent document and login information for the READI Web site. A copy of this Web page can be found in Appendix A. It is important to note, the research Web page instructed each faculty member to complete the survey questions based on their experience teaching online only.

A unique and helpful characteristic of the READI assessment is the ability for clients to add institution specific questions. Therefore, both the READI designed assessment tools and the NSOPF satisfaction questions were completed within the same tool. The results were hosted on the READI Web site. The research Web site instructed each individual to choose an anonymous alias and use it to populate the “name” field of the assessment. This alias was used as each respondent’s sole identifier and allowed for complete anonymity and organization for statistical correlation and analysis. The research was conducted during the fall 2008 semester.

Sample

At the time of data collection, there were approximately 230 online instructors teaching for the college. Each online faculty member was invited to participate via email; however, it was very unlikely that every online faculty member would participate. The desired participation level for this study was 45%, or approximately 100 individual online faculty members.

The sample size goal of 100 respondents was accomplished. The total sample size for the study was 110. However, several respondents did not complete the entire study. For the purposes of statistical analysis, missing data were deleted pair-wise.
The results in Table 1 reveal the demographics of the sample. Of the total respondents \((n=110)\), 41 were male (approximately 37%), and 69 were female (approximately 63%). Gender was not considered when testing the research questions.

### Table 1. Demographic Results of Sample

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>Male</td>
<td>41</td>
<td>37.3</td>
<td>37.3</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>69</td>
<td>62.7</td>
<td>62.7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>110</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**READI assessment**

The READI Assessment (now called SmarterMeasure) is a tool designed to determine student readiness for online learning. As a current customer of READI, this institution had full access to its use, implementation, and prior research. The institution had already been set up with an online account with READI; administrators are able to direct subjects to the specific institution’s portal of the Web site for online administering of the tool. The tool is designed to assess online learner readiness via six components: learning styles, individual attributes, technical competency, technical knowledge, on-screen reading speed and comprehension, and typing speed and accuracy (READI, 2009). Although the assessment contains six variables, this study focused on the findings related to learning styles only.

**Learning styles**

The learning styles portion of the READI (2009) was “adapted from the larger (70 item) learning styles instrument administered by http://www.memletics.com. The instrument was reduced to 35 items in READI and some of the wording was modified to apply to online courses.” The instrument embedded within the tool is based on the multiple intelligences theory of Howard Gardner. READI breaks the learning styles down as follows:

3. Verbal (linguistic): You prefer using words, both in speech and writing.
4. Physical (kinesthetic): You prefer using your body, hands and sense of touch.
5. Logical (mathematical): You prefer using logic, reasoning and systems.
6. Social (interpersonal): You prefer to learn in groups or with other people.

Working within Gardner’s multiple intelligences theory, the idea that individuals possess dominant learning preferences offers deep implications for the use of educational technologies. For example, individuals who are physical (kinesthetic) learners are presumably outside of their learning preference type when teaching an online class. Visual (spatial) learners would potentially find themselves within their preference type when teaching most online courses. Of course, class type, the kinds of instructional methods used and the extent to which online interaction is used will also affect these assumptions.

The questions from which the results are drawn ask questions related to tendencies and preference in learning venues. The results from the learning styles portion of the assessment are presented in graph format so test takers can view not only their dominant learning styles, but also those second, third, and so on. This study only took into consideration the dominant learning style of each individual.

**READI validity**

In 2007, READI commissioned Atanda Research to conduct a correlational study between READI scores and academic success. They found “forty-two statistically significant correlations between READI variables and measures of academic success and goodness of fit” (READI, 2009).
READI reliability

To verify the consistency of the measurements used in the READI Assessment, in 2008 the Applied Measurement Associates of Tuscaloosa, Alabama, conducted an item reliability test using the Cronbach alpha coefficient. For questions related to technical usage, learning styles and personal attributes they found Cronbach alpha coefficients ranging from .76 to .85. Overall, the reliability results revealed acceptable levels of measurement consistency and reliability.

NSOPF instrument

As presented in the literature review of this study, the bulk of faculty job satisfaction research hinges upon the findings of the NSOPF conducted by the Department of Education. Generally, researchers incorporate the data available from the survey results into their own studies. Because this study was concerned with satisfaction among distance learning faculty only, using the preexisting data from the national survey would have been useless.

The NSOPF contains eight job satisfaction questions. Each of the questions requires one of four responses to the question, “With regard to your job . . . , would you say you are very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with . . . .” The eight questions are as follows:

1. . . . the authority you had to make decisions about content and methods in your institutional activities?
2. . . . the institutional support for implementing technology-based instructional activities?
3. . . . quality of equipment and facilities available for classroom instruction?
4. . . . institutional support for teaching improvement (including grants, release time, and professional development funds)?
5. . . . your workload?
6. . . . your salary?
7. . . . the benefits available to you?
8. . . . your job at this institution, overall?

Because this study was solely concerned with online teaching satisfaction, the respondents were asked to answer the questions with regard to their online teaching only. Most of the online instructors teach both online and face-to-face.

NSOPF validity

The methodology report for the 2004 NSOPF contains information related to the validity of the instrument. Validity results were reported based on nonresponse bias analysis methods. This analysis reported the “difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate, using the design weights prior to nonresponse adjustment” (Heuer et al., 2005, p. I-3). The results revealed only a 4.3% nonresponse bias for faculty reporting from institutions granting only associates degrees.

Findings

Out of 110 valid responses, approximately 69% of respondents reported that they were very satisfied with the authority they had to make decisions about content and methods in institutional activities while 30% reported that they were somewhat satisfied. Less than 1% reported any level of dissatisfaction.

Approximately 56% of respondents reported that they were very satisfied with institutional support for implementing technology-based instructional activities while approximately 33% reported that they were somewhat satisfied. Less than 11% reported any level of dissatisfaction.

Approximately 36% of respondents reported that they were very satisfied with the quality of equipment and facilities available for classroom instruction while 50% reported that they were somewhat satisfied. Less than 14% reported any level of dissatisfaction.
Approximately 22% of respondents reported that they were very satisfied with institutional support for teaching improvement while approximately 48% reported that they were somewhat satisfied. Thirty percent reported some level of dissatisfaction.

Approximately 46% of respondents reported that they were very satisfied with their workload while approximately 31% reported that they were somewhat satisfied. Less than 23% reported any level of dissatisfaction.

Approximately 26% of respondents reported that they were very satisfied with their salary while approximately 46% reported that they were somewhat satisfied. Less than 28% reported any level of dissatisfaction.

Approximately 42% of respondents reported that they were very satisfied with their benefits while approximately 48% reported that they were somewhat satisfied. Ten percent reported some level of dissatisfaction.

Approximately 61% of respondents reported that they were very satisfied overall while approximately 34% reported that they were somewhat satisfied. Less than 6% reported any level of dissatisfaction.

**Learning styles frequencies**

Out of a total number of 110 respondents, 103 completed the learning style section. Out of a valid 103 respondents, approximately 28% were social learners while approximately 24% were verbal learners. Physical learners made up less than 5% of the sample.

**Results**

There was not a statistically significant difference in online instructor satisfaction with the authority to make decisions about content and methods based on individual instructor learning styles.

There was not a statistically significant difference in online instructor satisfaction with technology-based institutional support based on individual instructor learning styles.

The test for homogeneity of variance for satisfaction with equipment and facilities resulted in a non-significant value ($p > .05$). Therefore the variances were assumed equal and analysis of variance was conducted.

As presented in Table 2, analysis of variance testing resulted in a significant value ($p < .05$). The variance in satisfaction with decisions was explained by learning style. Therefore, the null hypothesis was rejected. A post-hoc test was necessary to further understand the nature of the statistically significant result.

<table>
<thead>
<tr>
<th>(I) Learning Style</th>
<th>(J) Learning Style</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Interval Lower Bound</th>
<th>95% Interval Upper Bound</th>
<th>Confidence Lower Bound</th>
<th>Confidence Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aural</td>
<td>Visual</td>
<td>- .92424</td>
<td>.37464</td>
<td>.183</td>
<td>-2.0526</td>
<td>.2041</td>
<td>-2.0526</td>
<td>.2041</td>
</tr>
<tr>
<td></td>
<td>Verbal</td>
<td>- .94872</td>
<td>.33433</td>
<td>.078</td>
<td>-1.9556</td>
<td>.0582</td>
<td>-1.9556</td>
<td>.0582</td>
</tr>
</tbody>
</table>

The results of the Tukey post hoc test in Table 3 further illustrate the nature of the statistically significant result. Respondents categorized as aural learners reported different levels of satisfaction with equipment and facilities than both physical and social learners.
<table>
<thead>
<tr>
<th></th>
<th>Physical</th>
<th>Logical</th>
<th>Social</th>
<th>Solitary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.43333(*)</td>
<td>.44699</td>
<td>-1.25269(*)</td>
<td>-.91667</td>
</tr>
<tr>
<td></td>
<td>.44699</td>
<td>.36909</td>
<td>.32924</td>
<td>.36909</td>
</tr>
<tr>
<td></td>
<td>.029</td>
<td>.061</td>
<td>.005</td>
<td>.177</td>
</tr>
<tr>
<td></td>
<td>-2.7796</td>
<td>-2.1950</td>
<td>-2.2443</td>
<td>-2.0283</td>
</tr>
<tr>
<td></td>
<td>-.0871</td>
<td>.0283</td>
<td>-.2611</td>
<td>.1950</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level.

The means plot displayed in Figure 1 reveals that aural learners \((n=6)\) reported significantly lower levels of satisfaction with the quality of equipment and facilities than both physical \((n=5)\) and social learners \((n=31)\).

![Means plot of satisfaction with equipment and facilities based on learning styles](image)

**Figure 1.** Means plot of satisfaction with equipment and facilities based on learning styles

**Discussion**

Although the null hypothesis was rejected for the research question, the findings do not necessarily suggest that individuals with particular learning preferences are more or less satisfied with teaching online. However, the findings do suggest that individuals with particular learning preferences are more satisfied with the quality of equipment and facilities available to them. Aural learners reported significantly lower levels of satisfaction with the quality of equipment and facilities than both physical and social learners. Aural learners rely heavily on sound to stimulate their learning; they learn well in face-to-face settings or through audio presentations. Physical learners learn by doing; generally they are not worried about breaking or damaging equipment or technology when they learn how to use it. Therefore, physical learners do not usually need as much guidance in how to manage equipment. Social learners learn well in groups or with other people; therefore, if they have any difficulty with a piece of equipment they are likely to seek help or guidance from a colleague or expert. Given this explanation, the statistically significant difference in the satisfaction with equipment and facilities between aural learners and both physical and social learners shines some light on the type of satisfaction in question. Based on these findings, it is not likely that the level of satisfaction was with the quality of equipment and/or facilities, but rather with the individual faculty member’s ability to manage them. Speculation regarding the quality of the equipment and facilities could not be made based on learning styles. However, it is possible that aural learners, while not understanding how to use a particular piece of equipment, perceive a lower level of quality than those who understand the tool in question. Therefore, given the nature of the research question and the results, it is likely that aural learners require more auditory guidance concerning the management, facilitation, and implementation of available equipment and facilities.

Working within Herzberg’s motivator-hygiene framework, equipment and facilities can be viewed as a mediator for aural learners. More specifically, for Herzberg, equipment and facilities would be categorized as hygienes, or demotivators. The reason for this categorization relies heavily on the nature of the aural learner, discussed above.
Further, this finding speaks to Pollicino’s (1996) theory where the onus of faculty satisfaction lies with the institution. In this case, for Pollicino, it is the institution’s responsibility to provide quality equipment and facilities and adequate training to effectively manage them. This connection contributed to some of the following recommendations.

There was no statistically significant difference in any of the other satisfaction question results based on individual faculty member learning preference. The data from the two satisfaction questions related to technology-based institutional support and teaching support were not homogenous. The results from the remaining five satisfaction questions resulted in non-significant differences based on online instructor learning preferences, although the variances in the data were homogenous. Based on these findings, there is no clear individual learning preference related to online instructor job satisfaction. It is also important to note, based on the extent to which satisfaction with work is related to intent to leave and/or perseverance in work, these findings were not congruent with Fuller et al. (2000).

**Recommendations for further practice**

Given the differing needs of aural learners, administrators, trainers and instructional designers can make educated decisions about how to support aural learners.

**Accommodating Aural Learners**

To accommodate the aural learners, institutions providing equipment and facilities to instructors should schedule face-to-face workshops or create audio/video tutorials for equipment and facilities training. Face-to-face workshops are fairly simple to organize and usually receive good responses from faculty. Although social learners did not report low levels of satisfaction in this area, these kinds of workshops serve their needs as well. Video tutorials discussing the use of available equipment are vital in the distance learning profession. Faculty members have busy schedules with teaching loads and committee obligations; they do not always have the time to seek out individual assistance with a piece of technology or equipment. Additionally, they do not always have the time to attend workshops. Using a user-friendly software program such as Adobe Captivate or Camtasia to create video and audio tutorials for the management of distance learning equipment and facilities is inexpensive and extremely beneficial. Not only do these kinds of tutorials provide flexible, asynchronous professional development opportunities for faculty, they serve the needs of the less satisfied aural learners as well.

**Provide Options**

Professionals in faculty development and online instructor training should provide multiple options for training and course development. Options in both training programs and online course development models allow instructors to make choices based on known or perceived strengths or preferences as well as challenge some to choose avenues of less comfort. This could also, according to Vroom’s (1964) expectancy theory, assist in individual instructor motivation to persevere through training, course development and online teaching. However, while it is important to provide a trainee-centered program, future online instructors should be encouraged to provide options for their students as well. Therefore, building on these findings and the work of Veronica and Lawrence (1997), instructors should be reminded to use methods and pedagogical tools outside of their preference to accommodate differences in student preference as well.

**Recommendations for further research**

The results from the two satisfaction questions related to technology-based institutional support and teaching support were not homogenous in variance. Therefore, future researchers should conduct this piece of the study again with the goal of receiving a larger sample size. With a larger sample size the results have a higher possibility of achieving homogeneity of variance.
The results of this study suggest that instructor learning preferences should not be taken into consideration when selecting individuals to teach in the online environment, if the purpose is to select individuals who will enjoy teaching online. However, future researchers should create a similar study as this one, except as the dependant variable incorporate online instruction quality instead of satisfaction. It is possible that while job satisfaction is generally not related to learning preference, quality of instruction may be. A future study such as this could link this research with Katzell’s et al. (1992) theory about work quality and satisfaction. Future researchers are encouraged to incorporate externally informed benchmarks for online teaching quality.

As the NSOPF is a product of the U.S. Department of Education, its application is designed for an American context. If the constructs upon which the NSOPF was developed are determined to have any obvious American biases, there is potential for this research to lose applicability in an international context. Therefore, further research related to the nature of faculty satisfaction in other cultural settings would be important.

First, future researchers should study the current state of knowledge related to faculty satisfaction to determine the applicability of the constructs contained in the NSOPF. If biases and/or discrepancies are found, researchers should seek out an existing tool designed to measure satisfaction in other cultural settings. Lastly, future research could determine, by way of a comprehensive review of existing international literature, broader constructs related to faculty satisfaction that would be applicable in multiple cultural settings.

This study did not take into consideration demographics such as race, gender or age. Further, this study did not take into consideration discipline affiliation, number of years taught online, employment status (full-time, adjunct, etc.), or highest degree earned. Each of these would be worthy variables to consider in future studies.

Next, further research relating to online instructor learning styles is necessary. As mentioned earlier, much of the research in the area of learning styles and preferences focuses on the student. A thorough study designed to determine whether instructors tend to teach toward their dominant learning style would be worthwhile. Further, one other worthwhile study would be to analyze the relationship between instructor learning styles and their aptitude and/or desire to learn and accept the utilization of a partucilare piece of instrucational technology or software (learning management system, Web 2.0 applications, etc.). For a theoretical framework, future researchers should analyze appropriateness of incorporating Fred Davis and Richard Bagozzi’s Technology Acceptance Model (TAM) as studied by Bagozzi and Warshaw (1992).

Lastly, the distribution of learning styles/preferences for instructors, found in this study, seem to be congruent with that of national averages for student preferences (Smarter Measure 2010). However, future studies relating instructor preference to student preference in the online environment should be pursued.

References


Yu-Chen Hsu¹, Hsin Ning Jessie Ho¹*, Chin-Chung Tsai¹, Gwo-Jen Hwang¹, Hui-Chun Chu², Chin-Yeh Wang³ and Nian-Shing Chen⁴

¹Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan, #43, Sec. 4, Keelung Rd., Taipei, 106, Taiwan // ²Department of Computer Science and Information Management, Soochow University, 56 Kueiyang Street, Section 1, Taipei, Taiwan 100, R.O.C. // ³Research Center for Science and Technology for Learning, National Central University, No.300, Jongda Rd., Jongli City, Taoyuan County 32001, Taiwan R.O.C. // ⁴Department of Information Management, National Sun Yat-sen University, No. 70, Lien Hai Rd., Kaohsiung, Taiwan, 80424, R.O.C. // hsuyuchen@mail.ntust.edu.tw // hojessie@gmail.com // cctsai@mail.ntust.edu.tw // gjhwang.academic@gmail.com // carolhcchu@gmail.com // chinyeawang@gmail.com // nschen@mis.nsysu.edu.tw

*Corresponding author

(Submitted April 18, 2011; Revised June 5, 2011; Accepted July 12, 2011)

ABSTRACT

This paper provides a content analysis of studies in technology-based learning (TBL) that were published in five Social Sciences Citation Index (SSCI) journals (i.e. the British Journal of Educational Technology, Computers & Education, Educational Technology Research & Development, Educational Technology & Society, the Journal of Computer Assisted Learning) from 2000 to 2009. A total of 2,976 articles were cross-analyzed by three categories including research topic, research sample group, and learning domain. It was found that “Pedagogical design and theories” was the most popular research topic, “Higher Education” was the most utilized sample group, and “Non-specified” and “Engineering/Computer sciences” were the most selected learning domains in the last decade. However, topics in “Motivation, Perceptions and Attitudes” drew more attention in the latest five years, while the number of articles in “Digital game and intelligent toy enhanced learning” and “Mobile and Ubiquitous Learning” grew significantly between 2005 and 2009. Furthermore, the Chi-square analysis results showed that there were significant associations among these three categories. The results of the analysis provide insights for educators and researchers into research trends and patterns of technology-based learning.

Keywords

Research trends, Technology-based learning, Content analysis

Introduction

Technology-based learning is increasingly important in the 21st century. However, studies in analyzing its appropriateness, its application and practices, and its influence on today’s education are still under debate and discussions (e.g. Chan, Hue, Chou, & Tzeng, 2001; Shih, Feng, & Tsai, 2008). Besides, numerous studies in this area have explored a wide range of topics, such as the improvement in technology-based learning environments, the effectiveness of web-based instruction, and the integration of new technology into classrooms. Therefore, reviewing the research trend of technology-based learning may help the researchers in related fields to identify their research interests and design considerations. Besides, the study will also provide policymakers with a reference to make plans in the future. Hence, it is important and meaningful to examine the technology-based learning research trends at this point.

This study intends to investigate the research trends of technology-based learning from 2000 to 2009. Five major research journals were selected for analysis, namely the British Journal of Educational Technology (BJET), Computers and Education (C&E), Educational Technology Research & Development (ETR&D), Educational Technology & Society (ET&S), and the Journal of Computer Assisted Learning (JCAL). The research questions addressed by this paper include:

1. What research topics related to technology-based learning were published in these selected journals from 2000 to
2. What research sample groups related to technology-based learning were selected in these published articles from 2000 to 2009? And how did the sample group selections shift between the first five years (2000-2004) and the second five years (2005-2009)?

3. What research learning domains related to technology-based learning were adopted in these published articles from 2000 to 2009? And how did the learning domain shift between the first five years (2000-2004) and the second five years (2005-2009)?

4. Is there any significant association between the research topic and the selection of the research sample group for these publications from 2000 to 2009?

5. Is there any significant association between the research topic and the adoption of the learning domain for these publications from 2000 to 2009?

Reviews of TBL research trends

The prevalent use of computing and communication technologies has increased in education and thus, learning is no longer limited to the traditional environment. Communication technologies such as the Internet, digital programs and systems, Personal Digital Assistants (PDA), and simulation games have been integrated into instruction to support learning. As a result, technology-based learning refers to the process of learning constituted with electronic technology (Cavus & Kanbul, 2010).

As more technological innovations are developing, their applications to education are also believed to be evolving. For example, Yengina, Karahocab A., Karahocab D., and Ozcinarc, (2010) have predicted that more technology-based learning (TBL) will occur with the newly developed devices or concepts, such as personalized and adaptive e-learning, portfolio collections, and more advanced online mind tools. In addition, the new web-based social networking tools such as Facebook, Twitter, YouTube, or wiki will also expect to become an integral component of TBL (e.g. Clark & Mayer, 2008). As a result, examining what previous TBL research has accomplished and emphasized may help researchers to identify the research trends or design criteria. For example, Waight, Willging, and Wentling (2004) clarified e-learning reports published by government, business, and associations to understand the impacts and the focus of e-learning in the United States from 1999 to 2003 and the identified trends were: lifelong learning, improvements in technology, demand for high level skill workers, pervasiveness of computers, globalization, new ways of learning from new technologies, and improvement of the learning quality via the technology. A similar study conducted by Shih, et al., (2008) was to investigate the research trends of cognitive studies in e-learning from 2001 to 2005. Five journals were selected for analysis (i.e. the Journal of Computers and Education, the British Journal of Educational Technology, Innovations in Education and Teaching International, Educational Technology Research & Development, and the Journal of Computer Assisted Learning.) that the first three ranked research topics were “Instructional Approaches,” “Learning Environment,” and “Metacognition”; the studies in “Instructional Approaches,” “Information Processing,” and “Motivation” received the highest citation counts, and the research data tended to be gathered from learners’ electronic log files or online-messages.

Furthermore, with a focus on a topic of interaction, Karatas (2008) examined three research journals (American Journal of Distance Education, Quarterly Review of Distance Education, & Distance Education) from 2003 to 2005 and concluded that (1) publications related to interaction were most frequently conducted in the years 2003 and 2004, (2) most of these articles were published in the journals of American Journal of Distance Education and Quarterly Review of Distance Education, (3) Articles in “Evaluation” of distance education was the most used topic. Like Karatas’s study, Zawacki, Bäcker, and Vogt (2009) selected 695 articles from five research journals (i.e. Open Learning, Distance Education, the American Journal of Distance Education, the Journal of Distance Education, the International Review of Research in Open and Distance Learning) to find the research trends in distance education from 2000 to 2008 based on the articles’ research areas, methods, and authorship patterns. The results revealed that research in distance education was dominated by studies that focus on interaction and communication patterns in computer-mediated communication, instructional design issues, learner characteristics, and educational technology.

Taken together, all of these studies utilized systematic content analysis methods and provided some insights into the research trends of TBL. This study, however, extends the previous studies’ frameworks to analyze the TBL research trends according to three categories, namely the major research topics, the research sample groups, and the learning
domains. Five research questions were raised in accordance with the aforementioned literatures and the panel discussions of four experts in educational technology. Besides, the current study selected major TBL journals and conducted a longitudinal analysis for a ten-year-period to present an all-inclusive overview. As it is difficult to include all of the relevant journals for analysis, four experts in educational technology were consulted to select the five major journals in this study. The selected journals were also similar to Shih, et al.’s study (2008) and the selected period which contains ten-year analysis should represent as a complete TBL research review for a decade. Moreover, the current research further analyzes the trend by examining the associations among the three identified categories and compares their changes between the first and the last five years. For instance, this study explores how the research topic may be related to the adoption of the research sample. Such a trend issue has not been carefully addressed by previous review papers. Overall, the current study contributes a more complete and comprehensive report of the TBL research trends compared to the previous studies.

**Methods**

This study examines papers relevant to TBL in the SSCI database from 2000 to 2009. The five selected journals were: *the British Journal of Educational Technology (BJET)*, *Computers and Education (C&E)*, *Educational Technology Research & Development (ETR&D)*, *Educational Technology & Society (ET&S)*, and *the Journal of Computer Assisted Learning (JCAL)*. These journals are widely accessed and recognized as high impact factors as released by the Institute for Scientific Information (ISI) Journal Citation Reports.

There were a total of 3,655 document items from 2000 to 2009 published by these five journals. Only papers that were identified as being of the type “articles” in the SSCI were analyzed. Publications such as “book reviews,” “letters,” and “editorial materials” were all excluded from this study. A further comprehensive review was then carried out. Finally, a total of 2,976 articles were selected for the analysis. In other words, this review only excluded publications which were not categorized as research “articles.” It included all of the remaining papers published in these journals without utilizing other filtering criteria. Hence, it is expected that this review can provide a more thorough view of TBL research.

**Research topics**

One of the major purposes of this study was to categorize research topics in TBL research to help identify the research trends from 2000-2009. After conducting several rounds of expert panel discussions, some important conferences and handbooks were selected for developing a framework of analyzing the research topics. The included conferences and handbooks were the 5th International Conference on e-Learning (ICEL 2010), the eLearning Forum 2010, the 10th IEEE International Conference on Advanced Learning Technologies (ICALT 2010), the Technology Enhanced Learning Conference 2009 (TELearn 2009), the 2010 International Conference on e-Education, e-Business, e-Management and e-Learning (IC4E 2010), the ICCE Conference on Technology, Pedagogy and Education (ICCE 2009), and the 5th International Conference on Open and Distance Learning (ICODL 2009), as well as Handbook: A Comprehensive Guide to Online Learning (Carliner & Shank, 2008), and the SAGE Handbook of E-learning Research (Andrews & Haythornthwaite, 2007). By integrating the conference themes and handbook chapters, the experts then formed a framework and concluded thirteen sub-categories for the research topics, as presented in Table 1.

<table>
<thead>
<tr>
<th>Table 1. The TBL research topic sub-categories and sample topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Topic Sub-category</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
</tbody>
</table>

356
<table>
<thead>
<tr>
<th>Section</th>
<th>Topic</th>
<th>Subtopics</th>
</tr>
</thead>
</table>
(2) User Prior Knowledge and Background Analysis  
(3) Innovative Pedagogical Models  
(4) Next Generation Learning Strategies  
(5) The Potential of Brain Science and its Impact on Learning |
(2) Critical Thinking and Problem Solving Development  
(3) E-Library and Learning Resources  
(4) Learners Diversity, Inclusiveness and Inequality  
(5) Supporting Students Experience |
| 5.      | Artificial Intelligence in Education | (1) Intelligent Educational Systems  
(2) Ambient Intelligence and Smart Environments for Learning  
(3) Semantic Web and Ontology for Learning Systems  
(4) Data Mining and Web Mining in Education |
| 6.      | Computer Supported Collaborative Learning | (1) Web 2.0 and Social Computing for Learning and Knowledge Sharing  
(2) Collaborative and Group Learning  
(3) On-line Dictionaries and Language development  
(4) Collaborative and Social Technologies |
| 7.      | Mobile and Ubiquitous Learning | (1) Wireless, Mobile and Ubiquitous Technologies for Learning  
(2) Pervasive Computing for Learning  
(3) U-Computing in Learning  
(4) Context-Aware U-Learning |
| 8.      | Digital Game and Intelligent Toy Enhanced Learning | (1) Gaming, Simulation and Virtual Worlds  
(2) Identity in Gaming to Learn: Roles and Role-Playing - Collaboration, Competition and Community  
(3) Non-Visual Senses (Smelling, Touching, Hearing)  
(4) Game-Based Learning/Training  
(5) Multiplayer and social game design  
(6) Virtual Characters, Virtual Storytelling and Game Narrative |
| 9.      | E-Assessment and New Assessment Theories and Methodologies | (1) E-Assessment Strategies  
(2) Automatic e-Assessment  
(3) Collaborative e-Assessment  
(4) E-portfolio Implementation and Impact |
| 10.     | Special Needs Education | (1) Technology enhanced Special Needs Education  
(2) Learning Disabilities  
(3) Technology Enhanced Learning for Students with Special Needs |
| 11.     | Motivation, Perceptions and Attitudes | (1) Technology Acceptance of Technologies in Education  
(2) Readiness of Technologies in Education  
(3) Sociological and Psychological Dimensions of Technology-Enhanced Learning  
(4) Self-regulation and Metacognition |
| 12.     | Learning Behaviors, Usage Patterns and Discourse Analysis | (1) Social Networking and Interactive Participation  
(2) Learning Culture  
(3) User Attitude and Technology Acceptance  
(4) Analysis of Asynchronous or Synchronous Interaction  
(5) Reflective Discourse |
(2) Organizational Policies and Strategies for Technology-Enhanced Learning Implementation  
(3) Global Trends in Technology-Enhanced Learning |
Research sample groups

After the four experts’ discussion and consensus, seven sub-categories for the research samples were identified. The finalized sub-categories were: 1. Elementary school, 2. Junior and Senior high school, 3. Higher education, 4. Teachers, 5. Adults, 6. Others, and 7. Non-specified.

Research learning domains

The same experts discussed the usage of the sub-categories for the learning domains. The learning domain for each paper included in this review was categorized into one of the following sub-categories: 1. Science (e.g., Physics, Chemistry, and Biology, Medical and Sport Science), 2. Mathematics, 3. Arts & Language, 4. Social Studies, 5. Engineering (including Computers), 6. Others, and 7. Non-specified.

Data coding and analysis

This study used all of the empirical research articles published in BJET, C&E, ETR&D, ET&S, and JCAL from 2000 to 2009 (ten years) to examine the research trends in the TBL field. Content analysis was utilized to classify the research topics, sample groups, and learning domains of the articles.

It is noted that many research articles addressed more than one research topic or sample group; therefore, the research topic and sample group category was coded into two levels. The primary matched sub-category was coded as “the first research topic” or “the first sample group” and the preceding matched sub-category was considered as “the second research-topic” or “the second sample group.” The matched sub-categories after the “second research-topic” or “the second sample group” were excluded. If the article had only one research topic or one sample group, the field of the second research topic or second sample group was coded “none.” For example, if an article aimed to evaluate whether a new learning system enhanced the learning motivation of fourth grade students, the article’s first research-topic would be coded “2” (i.e., Evaluation of Learning Systems, Platforms and Architectures) and the second research-topic would be coded “11” (i.e., Motivation, Perceptions and Attitudes).

Finally, a one-level coding process was involved in categorizing the learning domains, as based on our pilot coding for about 200 papers, most of the papers did not involve more than one learning domain. Articles with only one major learning domain were coded to the designated learning domain, and articles with more than one major learning domain or other than the aforementioned learning domain sub-categories were accounted as “Others.” However, only around 5% of the papers were coded as “Others” for the learning domains (presented later).

The coding process was undertaken manually by the researchers. Seven doctoral students majoring in educational technology helped to categorize the articles based on the aforementioned categories (research topics, research sample groups, and learning domains). Two post-doctoral researchers then randomly selected 401 articles and followed the same coding process to compare the coding results by research topic. The results reached an agreement of 0.88. Since the research sample group and learning domain information was less ambiguous in most of the articles, the researchers only randomly located 100 articles to compare the coding results. The results of both categories reached the same agreement of 0.90.

After the classification, the researchers proceeded to assign weighted scores to each article. First, a score of 1 was given to “Regular articles” (n = 2,806) and a score of 0.5 was applied to “Short articles” (that is, the “Colloquium” articles in BJET, n = 170). The total weighted score was 2,891. Then, a score of 0.6 was assigned to the first research-topic and a score of 0.4 was allocated to the second research-topic when an article covered two matched research topic sub-categories. If the article solely had one matched research topic, this topic received a total score of 1 (for the “short articles”, the score of the first research topic was 0.3 and the score of the second research topic was
Similarly, a score of 0.6 was assigned to the first sample group and a score of 0.4 was allocated to the second one when an article included two matched sample group sub-categories. If the article had only one matched sample group, the identified sample group received a total score of 1 (for the “short articles”, a score of 0.3 was applied to the first sample group while a score of 0.2 was assigned to the second one. The articles with merely one sample group had a score of 0.5). Finally, since only one learning domain was assigned to each article, a score of 1 was given to articles based on the specified learning domain (for the “short articles”, a score of 0.5 was applied to the designated learning domain.)

After the initial coding process, a descriptive statistical analysis was then employed to report the data. Pearson’s Chi-Square analysis was also used to investigate if there were significant associations among these three categories. For the categories of research topics and research samples, the Chi-Square analysis was conducted based on the first matched sub-category. A cell number of less than or equal to 5.0 was excluded for analysis. A post hoc test in obtaining the adjusted residual values was then utilized to examine the difference between the expected values and the actual values for the factors that contributed to the associations. That is, when the confidence interval was 95%, the adjusted residual values were compared to the standardized z-value (z = 1.96 or -1.96). If the absolute values of the adjusted residual values were larger than 1.96, the factors were considered to contribute to the associations.

Results

**Research question 1**: What research topics related to technology-based learning were published in these selected journals from 2000 to 2009? And what were the topic variations between the first five years (2000-2004) and the second five years (2005-2009)?

Based on the aforementioned score allocation, the score under each sub-category of research topic is presented in Figure 1. From 2000 to 2004, the most published research topic was “Pedagogical Design and Theories” (249.4), followed by “Policies, social culture impacts and trends for technology-enhanced learning” (153.9), and “Development of New Learning Systems, Platforms, and Architectures” (138.3). The least published research topic was “Digital Game and Intelligent Toy Enhanced Learning” (8.8). On the other hand, from 2005 to 2009, the most published research topic was “Pedagogical Design and Theories” (501.7), followed by “Motivation, Perceptions and Attitudes” (240.8), and “Development of New Learning Systems, Platforms, and Architectures” (176). The least published research topic sub-category was “Special Needs Education” (19.4). Furthermore, only the number of articles in “Policies, social culture impacts and trends for technology-enhanced learning” reduced from 153.9 to 120.9 between the two periods.

The results of the Pearson’s Chi-Square analysis also revealed that the published research topics were significantly different between the initial five years (2000-2004) and the latest five years (2005-2009) ($\chi^2 = 89.52$, $p < 0.05$). The major difference was that the “Policies, social culture impacts and trends for technology-enhanced learning” research sub-category showed a declining trend (from 14.13% to 6.71%), while researchers demonstrated more interest in “Digital game and intelligent toy enhanced learning” topics (from 0.81% to 3.82%) between the two periods. Furthermore, the issues in “Pedagogical Design and Theories” increased from 22.90% to 27.84% and the topics in “Mobile and Ubiquitous Learning” also grew from 1.40% to 2.59% between the two periods. However, the research in “Development of New Learning Systems, Platforms and Architectures” dropped from 12.7% to 9.77%, and the topics in “Evaluation of Learning Systems, Platforms, and Architectures” also reduced from 10.13% to 7.99% between the two periods. In sum, the more recent TBL researchers were inclined to focus on personalized learning experiences by integrating the latest innovations into pedagogical practices (e.g. game-based learning, m-learning, and u-learning). Nevertheless, they expressed less interest in general TBL issues such as TBL policy impacts or learning system build-up than in the first period (2000-2004).
Research question 2: What research sample groups related to technology-based learning were selected in these published articles from 2000 to 2009? And how did the sample group selections shift between the first five years (2000-2004) and the second five years (2005-2009)?

This paper analyzed the publications by research samples and the results are displayed in Figure 2. From 2000 to 2004, research samples in “Higher Education” were utilized most (399.8), followed by “Non-specified” (301.1), “Junior and Senior High School” (104.4), “Elementary School” (98.1), and “Teachers” (90.2). The least employed research sample was “Others” (13.0). From 2005 to 2009, research samples in “Higher Education” were still used for most of the TBL research papers (830.1), followed by “Non-specified” (269.0), “Junior and Senior High School” (215.2), “Teachers” (199.7), and “Elementary School” (171). However, the number of articles in the “Non-specified” group reduced from 301.1 to 269 between the two periods.

The results of the Pearson’s Chi-Square analysis also showed that the research sample groups were significantly different between the initial five years (2000-2004) and the more recent five years (2005-2009) ($\chi^2 = 84.24$, p < 0.05). The major difference was that the sample group “Non-specified” declined from 27.65% to 14.93% and the
recruited samples in “Higher Education” increased from 36.71% to 46.07% between the two periods. Furthermore, this study also revealed that the trend of using “Junior and Senior High School” and “Teachers” sample groups increased from 9.59% to 11.94% and from 8.28% to 11.08%, respectively. On the other hand, the percentage of the studies on “Adults/Working Adults” dropped from 7.57% to 5.29% between the two periods.

![Figure 2. Scores and trends of TBL research sample group selections from 2000 to 2009](image)

**Research question 3:** What research learning domains related to technology-based learning were adopted in these published articles from 2000 to 2009? And how did the learning domain shift between the first five years (2000-2004) and the second five years (2005-2009)?

This paper analyzed the publications by the learning domains involved in TBL research, and the results are displayed in Figure 3. As found in Figure 3, the results of the content analysis showed that from 2000 to 2004, the “Non-specified” learning domain was found in most of the TBL publications (n = 501), followed by “Social Studies” (n = 153), “Engineering” (n = 136), “Science” (n = 130), “Arts & Languages” (n = 76), “Others” (n = 47), and “Math” (n = 46). From 2005 to 2009, the number one ranked learning domain was still “Non-specified” (n = 644.5), followed by “Engineering” (n = 378.5), “Science” (n = 218), “Social Studies” (n = 211), “Arts & Languages” (n = 133.5), “Others” (n = 109), and “Math” (n = 107.5).

The results of the Pearson’s Chi-Square analysis also reported that the adoption of learning domain was significantly different between the initial five years (2000-2004) and the later five years (2005-2009) ($\chi^2 = 56.42, p < 0.05$). The major difference was that the research trend in “Engineering” increased from 12.49% to 21.0% while the “Non-specified” learning domain decreased from 46.01% to 35.77%, even though the number of published articles in the “Non-specified” learning domain increased between these two periods (from 501 to 644.5). Additionally, studies in “Math” and “Others” also had more growth between the two periods (from 4.22% to 5.97% and from 4.32% to 6.05%, respectively).
Figure 3. Scores and trends of TBL research learning domain utilization from 2000 to 2009

Research question 4: Is there any significant association between the research topic and the selection of the research sample group for these publications from 2000 to 2009?

The findings of the content analysis showed that most of the published articles with different research topics preferred to utilize research sample groups in “Higher Education” (n = 1269). Pearson’s Chi-square analysis was then used to find the associations between the research topics and the research sample groups. The analysis was made based on the first matched sub-categories.

The results revealed that in the last decade (from 2000 to 2009), the research topics were significantly associated with the selections of the sample groups ($\chi^2 = 448.30, p < 0.05$). The post-hoc test results of obtaining residuals were used to explain the possible association. The Chi-square cells with significant adjusted residuals are presented in Table 2. As shown in Table 2, the adjusted residual values indicated that the topics in “Policies, social culture impacts and trends for technology-enhanced learning” with the “Non-specified” sample group had the fastest growing trend (residual = 13.9), followed by the “Motivation, Perceptions, and Attitudes” with the “Teachers” sample group (residual = 6.1), and the “Digital Game and Intelligent Toy Enhanced Learning” with the “Elementary School” sample group (residual=5.6). Besides, studies concerning “Special Needs Education” with the “Adults/Working Adults” sample group increased in the past ten years (residual = 3.3). The research topics in “E-assessment and New Assessment Theories and Methodologies” with the “Junior and Senior High school” sample group also showed climbing trends. Similarly, the research issues in “Computer Supported Collaborative Learning”, “Learning Behaviors, Usage Patterns and Discourse Analysis” and “Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competencies Management” with the “Higher Education” sample group (residuals = 4.5, 3.1, 2.0, respectively) and the research issues in the “Mobile and Ubiquitous Learning” with the “Elementary School” sample group (residual = 2.9) were explored more frequently over the last ten years.

On the other hand, research in “Motivation, Perceptions, and Attitudes” with the “Non-specified” sample group (residual = -8.2) showed a declining trend, followed by the “Policies, social culture impacts and trends for technology-enhanced learning” with the “Higher Education” sample group (residual = -7.8), and the “Computer Supported Collaborative Learning” with the “Non-specified” sample group (residual = -4.9). Besides, the topics in “Digital Game and Intelligent Toy Enhanced Learning” and “Artificial Intelligence in Education” with the “Higher Education” sample group (residuals = -3.5, -2.2, respectively), the issues of “Evaluation of Learning Systems, Platforms, and Architectures” with the “Teachers” sample group (residual = -2.4), and “Learning Behaviors, Usage Patterns and Discourse Analysis” with the “Non-specified” sample group (residual = -2.8) implied a decreasing tendency for the last decade.
### Table 2. Significant associations between research topics and research sample groups from Chi-square analysis (2000-2009)

<table>
<thead>
<tr>
<th>Research Topics</th>
<th>Research Sample Groups</th>
</tr>
</thead>
</table>
| 2. Evaluation of Learning Systems, Platforms and Architectures                   | Teachers (AR = -2.4)
   (n=13)<sup>b</sup> Higher Education (AR = 2.0)(n=50) |
| 5. Artificial Intelligence in Education                                         |                                                                            |
| 6. Computer Supported Collaborative Learning                                    |                                                                            |
| 7. Mobile and Ubiquitous Learning                                              |                                                                            |
| 8. Digital Game and Intelligent Toy Enhanced Learning                           |                                                                            |
| 9. E-Assessment and New Assessment Theories and Methodologies                  |                                                                            |
| 10. Special Needs Education                                                    |                                                                            |
| 11. Motivation, Perceptions and Attitudes                                       |                                                                            |
| 12. Learning Behaviors, Usage Patterns and Discourse                           |                                                                            |

Notes: Pearson’s Chi-Square = 448.30, degrees of freedom = 72, \( p < 0.05 \)

<sup>a</sup> (AR): Adjusted residual values (AR with absolute values larger than 1.96 are presented.)

<sup>b</sup> (n): Total number of articles

### Research trends for the first five years (2000-2004) and the latest five years (2005-2009)

To better understand if the research topics were significantly associated with the selections of sample groups between the two five-year-periods, Pearson’s Chi-Square analysis was conducted again to examine the association by the two periods, respectively. The Chi-square test results supported that the research topics were significantly associated with the selections of sample groups for the two respective periods (\( \chi^2 = 203.98, 294.99 \), respectively, \( p < 0.05 \)). A post-hoc test was also conducted to get the adjusted residual values for further analysis, as shown in Table 3.

### Table 3. Significant associations between research topics and research sample groups from Chi-square analysis in the two five-year-periods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Evaluation of Learning Systems, Platforms and Architectures</td>
<td>1. Adults ( (AR = 2.1)^a ) (n = 13)&lt;sup&gt;b&lt;/sup&gt; 2. Teachers ( (AR = -2.6) ) (n = 6)</td>
<td></td>
</tr>
<tr>
<td>3. Pedagogical Design and Theories</td>
<td>1. Higher Education ( (AR = -2.1) ) (n = 220) 2. Non-specified ( (AR = 2.4) ) (n = 100)</td>
<td>Higher Education ( (AR = 2.9) ) (n = 36)</td>
</tr>
<tr>
<td>5. Artificial Intelligence in Education</td>
<td>1. Higher Education</td>
<td></td>
</tr>
</tbody>
</table>
As presented in Table 3, the results indicated that the number of papers in “Computer Supported Collaborative Learning” and “Learning Behaviors, Usage Patterns and Discourse Analysis” with the “Higher Education” sample group maintained an increasing trend from the first period to the second period (residual values all larger than 2). Similarly, the published articles in “Policies, social culture impacts and trends for technology-enhanced learning” with the “Non-specified” sample group maintained a growing trend for the two periods (residuals were 8.2 and 10.3, respectively). Additionally, although the issues in “Motivation, Perceptions, and Attitudes” gained more attention during the second period, the used sample group changed from “Higher Education” to “Teachers” (residuals = 3.0, 5.7, respectively). Besides, in the first period, the research topics in “Artificial Intelligence in Education” showed a stronger association with the “Non-specified” sample group, while issues concerning “Mobile and Ubiquitous Learning” were examined mainly with the “Elementary Education” sample group (residuals = 3.1, 2.9, respectively). In the recent five years (second period), the topics in “Evaluation of Learning Systems, Platforms, and Architectures” gradually emphasized the “Adults/Working Adults” sample group (residual = 2.1) while the “Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competencies Management” issues were mainly investigated with the “Higher Education” sample group (residual = -2.1). Finally, the publications related to “Digital Game and Intelligent Toy Enhanced Learning” also started to proliferate in the “Elementary School” field in the past five years (residual = 5.0).

Conversely, the topics in the area of “Policies, social culture impacts and trends for technology-enhanced learning” with the “Higher Education” sample group (residuals = -5.9 and -4.4, respectively) and the sub-category of “Computer Supported Collaborative Learning” with the “Non-specified” sample group revealed a declining trend (residuals from -3.3 to -3.7, respectively) for the two periods. Overall, it is interesting to note that from 2000-2004, the topics of “Motivation, Perceptions, and Attitudes” were mainly conducted with the “Higher education” sample group, and no significant number of articles about “Digital Game and Intelligent Toy Enhanced Learning” with the “Elementary school” group was included. Therefore, the research trends in selecting the sample groups for these
topics have altered in the recent five years.

Research question 5: Is there any significant association between the research topic and the adoption of the learning domain for these publications from 2000 to 2009?

The results of the content analysis showed that “Non-specified” was the most commonly used learning domain for the published articles with different research topics in the past ten years ($n = 1198$). Besides, Pearson’s Chi-square analysis was used to find the associations between the research topics and the research sample groups. Again, the data was analyzed only based on the “first” matched sub-categories, and the results are displayed in Table 4. The results indicate that during the last decade, the research topics were significantly associated with the learning domain selections ($\chi^2 = 368.46$, $p < 0.05$). A post-hoc test was also administered to acquire the adjusted residual values for more detailed analysis.

The adjusted residual values revealed that the topics in “Policies, social culture impacts and trends for technology-enhanced learning” had the highest proportion in the “Non-specified” learning domains (residual = 13.2), followed by “Computer Supported Collaborative Learning” in “Social Studies” (residual = 5.4), and “Development of New Learning Systems, Platforms, and Architectures” in “Arts & Languages” (residual = 4.8). Besides, the topics of “Artificial Intelligence in Education” and “Pedagogical Design and Theories” showed a growing trend in the “Mathematics” learning domain (residuals = 4.5, 2.9, respectively). Similarly, the research issues in “Mobile and Ubiquitous Learning” were addressed more in the “Arts & Language” domain (residual = 2.8) and the topics in “Motivation, Perceptions, and Attitudes” and “Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competencies Management” in the “Engineering” learning domain gained much attention from researchers over the last ten years (residuals = 2.1, 4.4, respectively).

On the other hand, the research in “Computer Supported Collaborative Learning” with the “Non-specified” (residual = -5.7) learning domain, “Pedagogical Design and Theories” with the “Engineering” learning domain (residual = -2.5), and “Development of New Learning Systems, Platforms, and Architectures” with the “Non-specified” learning domain (residual = -2.3) were explored less frequently during the last decade.

Table 4. The significant associations between research topics and learning-domains from Chi-square analysis (2000-2009)

<table>
<thead>
<tr>
<th>Research Topics</th>
<th>Learning Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Development of Learning Systems, Platforms and Architectures</td>
<td>1. Arts &amp; Languages (AR = 4.8) ($n = 47$)</td>
</tr>
<tr>
<td></td>
<td>2. Non-specified (AR = -2.6) ($n = 119$)</td>
</tr>
<tr>
<td>3. Pedagogical Design and Theories</td>
<td>1. Math (AR = 2.9) ($n = 56$)</td>
</tr>
<tr>
<td></td>
<td>2. Science (AR = 2.6) ($n = 114$)</td>
</tr>
<tr>
<td></td>
<td>3. Others (AR = 2.4) ($n = 54$)</td>
</tr>
<tr>
<td></td>
<td>4. Engineering (AR = -2.5) ($n = 115$)</td>
</tr>
<tr>
<td></td>
<td>5. Non-specified (AR = -2.3) ($n = 289$)</td>
</tr>
<tr>
<td>4. Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competences Management</td>
<td>Engineering (AR = 4.4) ($n = 33$)</td>
</tr>
<tr>
<td>5. Artificial Intelligence in Education</td>
<td>Math (AR = 4.5) ($n = 14$)</td>
</tr>
<tr>
<td>6. Computer Supported Collaborative Learning</td>
<td>1. Social Studies (AR = 5.4) ($n = 60$)</td>
</tr>
<tr>
<td></td>
<td>2. Non-specified (AR = -5.7) ($n = 62$)</td>
</tr>
<tr>
<td>7. Mobile and Ubiquitous Learning</td>
<td>Arts &amp; Languages (AR = 2.8) ($n = 12$)</td>
</tr>
<tr>
<td>8. Digital Game and Intelligent Toy Enhanced Learning</td>
<td>Others (AR = 2.3) ($n = 10$)</td>
</tr>
<tr>
<td>11. Motivation, Perceptions and Attitudes</td>
<td>1. Engineering (AR = 2.1) ($n = 74$)</td>
</tr>
<tr>
<td></td>
<td>2. Math (AR = -2.0) ($n = 10$)</td>
</tr>
<tr>
<td></td>
<td>2. Science (AR = -4.8) ($n = 9$)</td>
</tr>
<tr>
<td></td>
<td>3. Engineering (AR = -4.3) ($n = 24$)</td>
</tr>
<tr>
<td></td>
<td>4. Social Studies (AR = -2.6) ($n = 22$)</td>
</tr>
</tbody>
</table>

Notes. Pearson’s Chi-Square = 368.46, degrees of freedom = 72, $p < 0.05$

a (AR): Adjusted residual values (AR with absolute values larger than 1.96 are presented.)
Similarly, Pearson’s Chi-Square analysis was conducted to investigate if the research topics were significantly associated with the adoption of the learning domain between the two five-year-periods. Again, the Chi-square results supported that the research topics were significantly associated with the selection of learning domain between the two periods ($\chi^2 = 182.77, 271.83$, respectively, $p < 0.05$). A post-hoc test was also conducted to obtain the adjusted residual values in order to illustrate the sources of the associations, as shown in Table 5.

### Table 5. The significant associations between research topics and learning domains from Chi-square analysis in the two five-year-periods

<table>
<thead>
<tr>
<th>Research Topics</th>
<th>Learning-domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Development of New Learning Systems, Platforms, and Architectures</td>
<td></td>
</tr>
<tr>
<td>1. Engineering ($AR = 2.2$) ($n = 28$)</td>
<td>1. Arts &amp; Languages ($AR = 4.6$) ($n = 30$)</td>
</tr>
<tr>
<td>2. Non-specified ($AR = -2.2$) ($n = 61$)</td>
<td>2. Non-specified ($AR = -2.0$) ($n = 58$)</td>
</tr>
<tr>
<td>2. Evaluation of Learning Systems, Platforms and Architectures</td>
<td></td>
</tr>
<tr>
<td>1. Arts &amp; Languages ($AR = 2.3$) ($n = 30$)</td>
<td></td>
</tr>
<tr>
<td>2. Non-specified ($AR = -2.0$) ($n = 58$)</td>
<td></td>
</tr>
<tr>
<td>3. Pedagogical Design and Theories</td>
<td></td>
</tr>
<tr>
<td>1. Science ($AR = 2.3$) ($n = 77$)</td>
<td>1. Social Studies ($AR = 3.0$) ($n = 24$)</td>
</tr>
<tr>
<td>2. Math ($AR = 2.3$) ($n = 41$)</td>
<td>2. Non-specified ($AR = -4.1$) ($n = 27$)</td>
</tr>
<tr>
<td>3. Engineering ($AR = -3.4$) ($n = 82$)</td>
<td></td>
</tr>
<tr>
<td>4. Non-specified ($AR = -2.5$) ($n = 104$)</td>
<td></td>
</tr>
<tr>
<td>1. Engineering ($AR = 4.9$) ($n = 26$)</td>
<td></td>
</tr>
<tr>
<td>5. Artificial Intelligence in Education</td>
<td></td>
</tr>
<tr>
<td>6. Computer Supported Collaborative Learning</td>
<td></td>
</tr>
<tr>
<td>1. Social Studies ($AR = 3.0$) ($n = 24$)</td>
<td>1. Social Studies ($AR = 4.5$) ($n = 36$)</td>
</tr>
<tr>
<td>2. Non-specified ($AR = -4.1$) ($n = 27$)</td>
<td>2. Non-specified ($AR = -4.1$) ($n = 35$)</td>
</tr>
<tr>
<td>7. Mobile and Ubiquitous Learning</td>
<td></td>
</tr>
<tr>
<td>12. Learning Behaviors, Usage Patterns and Discourse Analysis</td>
<td></td>
</tr>
<tr>
<td>1. Others ($AR = 3.3$) ($n = 16$)</td>
<td></td>
</tr>
<tr>
<td>1. Non-specified ($AR = 9.0$) ($n = 127$)</td>
<td>1. Others ($AR = 8.8$) ($n = 92$)</td>
</tr>
<tr>
<td>2. Social Studies ($AR = -2.3$) ($n = 13$)</td>
<td></td>
</tr>
<tr>
<td>3. Engineering ($AR = -3.6$) ($n = 6$)</td>
<td></td>
</tr>
<tr>
<td>12. Learning Behaviors, Usage Patterns and Discourse Analysis</td>
<td></td>
</tr>
<tr>
<td>1. Others ($AR = 3.3$) ($n = 16$)</td>
<td></td>
</tr>
<tr>
<td>1. Non-specified ($AR = 9.0$) ($n = 127$)</td>
<td></td>
</tr>
<tr>
<td>2. Social Studies ($AR = -2.3$) ($n = 13$)</td>
<td></td>
</tr>
<tr>
<td>3. Engineering ($AR = -3.6$) ($n = 6$)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 2001-2004 Pearson’s Chi-Square = 182.77, degrees of freedom = 72, $p < 0.05$
2005-2009 Pearson’s Chi-Square = 271.83, degrees of freedom = 72, $p < 0.05$

$^a$ (AR): Adjusted residual values (AR with absolute values larger than 1.96 are presented.)

$^b$ (n): Total number of articles
As presented in Table 5, the proportion of papers in “Policies, social culture impacts and trends for technology-enhanced learning” remained high in the “non-specified” learning domain from the first period to the second period (residual values = 9.0, 8.8, respectively). Additionally, the learning domain in the “Development of New Learning Systems, Platforms, and Architectures” topic changed from “Engineering” (first period) to “Arts & Languages” (second period) (residuals = 2.2, 4.6, respectively).

Besides, over the second period, the research topics in “Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competencies Management” attracted more attention in the “Engineering” learning domain (residual = 4.9), while “Mobile and Ubiquitous Learning” and “Evaluation of Learning Systems, Platforms, and Architectures” began to be addressed in the “Arts & Languages” field (residual = 3.6, 2.3). In addition, the topics in “Computer Supported Collaborative Learning” expressed an increasing tendency in “Social Studies” (residual = 4.5) while the sub-categories of “Pedagogical Design and Theories” and “Artificial Intelligence in Education” showed a gradually increasing tendency in the “Mathematics” learning domain (residuals = 2.3, 3.4, respectively). Furthermore, the researchers who were interested in the “Science” learning domain paid more attention to the topics in “Pedagogical Design and Theories” over the last five years (residual = 2.3).

Discussion

This study aims to explore the research trends for the TBL research articles based on their research topics, the sample group selections, and the learning-domain adoption. The results are discussed as follows:

The trends of TBL in research topics, sample selections, and learning domain adoption

The findings from content analysis show that the “Pedagogical Design and Theories” was the most popular research topic studied from 2000 to 2009 for the published TBL articles. Earlier TBL research was involved in debating whether the technology influenced learning (Clark, 2001; Kozma, 1991) and what factors might affect the instructions with various media representations (e.g., Mayer, 2001). Based on these valuable groundworks, many follow-up studies have suggested that pedagogical design is the major factor impacting learning in the TBL environment (Bernard, Abrami, Lou, & Borokhovski, 2004; Sitzmann, Kraiger, Stewart, & Wisher, 2006). As more relevant studies were apparent, the initial TBL theoretical frameworks and instructional design models have been established for enhancing and utilizing the TBL outcome (Koehler & Mishra, 2005; Clark & Mayer, 2008; O'Neil & Perez, 2006). It is expected that more research will examine the effectiveness of these pedagogical models for different TBL contexts in the future.

Additionally, the articles about “Motivation, Perceptions and Attitudes” have also increased between the two five-year-periods. Research about individual difference and preference toward different technologies have been studied since early twenty century and were limited in the traditional media such as texts, pictures, and video (Paivio, 1986; Salomon, 1984). After the prevalence of computers and online tools, related research extended their efforts to the advanced contexts such as animations, games, and Web-based learning. For example, learners' preferences or acceptance with respect to the online learning activities and their expression of persistence in the cyber learning environment were explored (Gan, 2005; Chen, Chen & Tsai, 2009; Yang & Tsai, 2008). Besides, a growing trend in studying the topics of “self-regulation” and “metacognition” should also be expected (Chu & Tsai, 2009) as more learner control functions are available from these latest advances than from traditional media.

The current study also shows that topics such as “Digital game and intelligent toy enhanced learning” have gained more attention in the recent five years. The result implies that students today are facing more advanced gaming experiences and educators started establishing a similar learning environment to help students maintain their attention on the learning tasks. However, how the pedagogical principles and the entertainment aspects can be well integrated will be a challenging issue to solve for future research (Harteveld, Guimarães, Mayer, & Bidarra, 2007).

By and large, the analyses of research topics by this paper (see Figure 1) suggest that in the last decade, the TBL research tended to develop more appropriate pedagogical and motivation models to enhance TBL learning. In addition, a small portion of research is increasing in finding whether learners' personalized learning experiences can be strengthened by integrating the latest innovations into pedagogical practices (e.g. m-learning, u-learning).
Compared to the publications in 2000-2004, the articles in 2005-2009 addressed less general TBL issues (e.g. TBL policy impacts or the development of wide-ranging learning systems) but more specific contextual applications (e.g. game-based learning).

Moreover, the present study also found that “higher education” is the major sample group employed the most in the TBL articles, while “Adults/Working adults” is the least employed. This finding implies that most of the TBL research is still being conducted by academic scholars. These experts have direct access to postsecondary students’ responses or performance results as research data and may have obstacles finding support from other institutes. However, as universal learning is emphasized in the US National Educational Technology Plan (US Department of Education, 2010), future research should put more efforts into adults or community learning outside the academic setting.

Finally, the findings reveal that the “Non-specified” learning domain was used most for the TBL studies in the past ten years. However, the percentage of articles in the “Non-Specified” learning domain has dropped in the recent five years. “Engineering” was the field which increased the most since 2005. These findings infer that the early TBL studies may have been reluctant to emphasize the research design based on domain-specific knowledge structure. However, the trend has changed such that more researchers have projected their TBL research in accordance with their particular interests in specific learning domains. The results also imply that more technical designers or engineers have started to cooperate with educators to include research-based learning theories into their application design, making the learning more effective.

**The cross analysis among research topics, sample, and learning-domain selections**

The current study indicates that “Higher Education” is the sample group utilized most in the TBL publications. However, the results of the Chi-square analysis show that the articles with topics in “Motivation, Perceptions, and Attitudes” and the “Teachers” sample group, “Digital Game and Intelligent Toy Enhanced Learning” and the “Elementary School” sample group increased the most in the nearest five years.

As revealed from the aforementioned literatures, most of the studies reviewing TBL trends from 1999 to 2008 concluded that the feature of communication innovations created by the new technologies played an important role in the TBL development (e.g. Waight et al, 2004; Karatas, 2008; Zawacki et al., 2009). The interaction methods and the design of instructional activities under the new TBL environment are different from traditional classroom environment that teachers and learners are required to adjust (Bernard et al., 2004). Therefore, the issues of “motivation, perceptions, and attitudes” were worth more investigation under TBL learning context. Early studies mainly focused on these topics from learners’ perspective (e.g. Northup, 2002) while the current finding reflects that the trend has been shifted to teachers in the recent five years (e.g. Mahdizadeh, Biemans, & Mulder, 2008). As more personalized communication advances become available, the motivation research in related to TBL learning is expected to be emphasized for other groups of learners, such as adult and special education.

Besides, the present result indicates that game-based learning has gained more attentions in the latest five years (2005-2009) and most of the relevant research was employed in elementary education. The result infers that children in the current era habitually play games in their daily life that they may be easily attracted by the game-related activities. Therefore, game-based learning may be more suitable for young children (e.g. Ke, 2008) and educators may hope to use game as “... a way to create relevant learning experiences that mirror students’ daily lives and the reality of their futures ” (US Department of Education, 2010, p. 9).

Furthermore, this study also reveals that the topics in “Policies, social culture impacts and trends for technology-enhanced learning” with the “Non-specified” learning domain, “Computer Supported Collaborative Learning” with “Social Studies”, and “Development of New Learning Systems, Platforms, and Architectures” with “Arts & Languages” have increased during the last decade. Additionally, the topics of “Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competencies Management” with “Engineering,” “Mobile and Ubiquitous Learning” and “Evaluation of Learning Systems, Platforms, and Architectures” with “Arts & Languages,” “Pedagogical Design and Theories” and “Artificial Intelligence in Education” with “Mathematics,” “Pedagogical Design and Theories” with “Science,” and “Learning Behaviors, Usage Patterns and Discourse Analysis” with “Others” have more growth in the recent five years. Clark and Salomon (1986) proposed that the
knowledge representations were associated with different domains or disciplines. As a result, different instructional methods embedded in the media presentation are necessary to be adapted for efficiently represent the information, which should compensate the learners’ cognitive representations. The assertion may help explain our current findings that the TBL researchers gradually target their research interests into specific learning domains for utilizing the instructional methods with the supported technologies. For instance, the computer-supported collaborative learning and group discussion may be applied as major instructional methods in the social studies learning domain; well-designed platforms or learning systems may be crucial to enhance arts and language learning. Moreover, these findings also infer that TBL implementation is not a domain-general task. In the last decade, studies in “Policies, social culture impacts and trends for technology-enhanced learning” seemed to provide essential guidelines or domain-general solutions for TBL practices. However, based on the current findings, it is predicted that adopting technologies in different domains will be a trend and also a challenge for the schools to face (Johnson, Levine, Smith, & Stone, 2010). Policy administrators, educators, or system developers need to carefully examine the specific TBL issues in different learning domains and provide the necessary support for future TBL development.

**Conclusion**

The present study examines the technology-based learning research trends between 2000 and 2009 from five major journals. Within this decade, TBL research trend has evolved from technology comparison to TBL-related pedagogical design theory development, from learners’ motivational issues to the teachers’, from Internet or learning system-based TBL to individualized and universal learning experiences, and from domain-general considerations to domain-specific advertency. The analysis shows that TBL context has become a common setting that educators are expected to deliver their instructions via technologies no matter to which learning groups or in which academic domains. Thus, policy makers and administrators should allocate more efforts and resources to develop better TBL implementation plans for the academy. For example, teacher education or training program should educate the teachers how the technologies should be used to enhance learners’ cognitive engagement and to represent good quality of TBL instructions. The recent development of TPACK (Technological Pedagogical Content Knowledge, Chai et al., 2010, 2011) model in helping teachers to design TBL-based curricular has started responding to such demand and inquiry.

Finally, this study excludes the relevant conference proceedings and other TBL related journals (e.g., IEEE publications) for analysis. It is suggested that future research should expand the data sources for more deliberate analysis. Additionally, this study is limited in including the relevant research in year 2010 for only providing a complete review of the TBL studies for a decade (2000-2009). Future research is encouraged to conduct similar studies with more current information and research data.

**Acknowledgements**

Funding for this research work is supported by the National Science Council, Taiwan, under grant numbers NSC 99-2631-S-011-001, 100-2911-I-011-502, and 100-2911-I-110-503.

**References**


Pedagogical Content Knowledge (TPACK) for meaningful learning with information and communication technology (ICT). *Computers & Education, 57*(1), 1184-1193.


