Educational Technology & Society
An International Journal

Aims and Scope

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

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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


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

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

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

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

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

Advisory board
Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Harley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasias 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; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoishi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan.

Executive peer-reviewers
http://www.ifets.info/

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.
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@iiiee.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@iiiee.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@iiiee.org (Subject: Submission for Educational Technology & Society journal).
Journal of Educational Technology & Society
Volume 16 Number 1 2013

Table of contents

Special issue articles

Guest Editorial: Innovative Technologies for the Seamless Integration of Formal and Informal Learning
Patrizia Ghislandi, Maria Grazia Ierardi, Tommaso Leo and Luca Spalazzi 1-3

Social Software and Educational Technology: Informal, Formal and Technical Values
Roberto Pereira, M. Cecilia C. Baranauska and Sergio Roberto P. da Silva 4-14

Beyond the School’s Boundaries: PoliCultura, a Large-Scale Digital Storytelling Initiative
Nicoletta Di Blas and Paolo Paolini 15-27

Investigating the Activities of Children toward a Smart Storytelling Toy
Nuri Kara, Cansu Cigdem Aydin and Kiersat Cagiltay 28-43

Integrating Traditional Learning and Games on Large Displays: An Experimental Study
Carmelo Ardito, Rosa Lanzilotti, Maria F. Costabile and Giuseppe Desolada 44-56

Affectibility in Educational Technologies: A Socio-Technical Perspective for Design
Elaine C. S. Hayashi and M. Cecilia C. Baranauska 57-68

Lecture Capture in Engineering Classes: Bridging Gaps and Enhancing Learning
Hasan Al Nashash and Cindy Gunn 69-78

Animating Civic Education: Developing a Knowledge Navigation System using Blogging and Topic Map Technology
Tien-Chi Huang and Chia-Chen Chen 79-92

Enriching Formal Language Learning with an Informal Social Component
Giuliana Dettori and Simone Torsani 93-103

PKS: An Ontology-based Learning Construct for Lifelong Learners
Flavio Manganello, Carla Falsetti, Luca Spalazzi and Tommaso Leo 104-117

Proposition and Organization of an Adaptive Learning Domain based on Fusion from the Web
Mohammed Chaoui and Mohamed Tayeb Laskri 118-132

Full length articles

Revisit the Effect of Teaching and Learning with Technology
Yuan-Hsuan Lee, Hersh Waxman, Jiun-Yu Wu, Georgette Michko and Grace Lin 133-146

Focus on Form in Live Chats
Wen-Chun Chen and Zohreh Eslami 147-158

Effect of a Virtual Chemistry Laboratory on Students' Achievement
Zeynep Tatli and Altinasa Ayas 159-170

Understanding Students' Competition Preference in Multiple-Mice Supported Classroom
Ben Chang, Min-Tun Chuang and Spes Ho 171-182

Developing a New Computer Game Attitude Scale for Taiwanese Early Adolescents
Eric Zhi-Feng Liu, Chun-Yi Lee and Jen-Huang Chen 183-193

Seamless Connection between Learning and Assessment- Applying Progressive Learning Tasks in Mobile Ecology Inquiry
Pi-Hsia Hung, Gwo-Jen Hwang, Yu-Fen Lin, Tsung-Hsun Wu and I-Hsiang Su 194-205

A Maturity Model for Assessing the Use of ICT in School Education
Mauricio Solar, Jorge Sabattin and Victor Parada 206-218

Student's Uncertainty Modeling through a Multimodal Sensor-Based Approach
Imène Jraidi and Claude Frasson 219-230
Exploring Factors Affecting Learner's Perception of Learning Information and Communication Technology: A HLM Analysis of a National Farmers' Training Program in Taiwan
Hsiu-Ping Yueh, Tzy-Ling Chen, Li-An Chiu and Wei-Chien Lin
231-242

Graphical Interaction Analysis Impact on Groups Collaborating through Blogs
Georgios Fessakis, Angelique Dimitracopoulou and Aggelos Palaiodimos
243-253

Can Speaking Activities of Residents in a Virtual World Make Difference to Their Self-Expression?
Soojeong Lee
254-262

Harnessing Collaborative Annotations on Online Formative Assessments
Jian-Wei Lin and Yuan-Cheng Lai
263-274

Effects of Game Design Patterns on Basic Life Support Training Content
Sebastian Kelle, Roland Klemke and Marcus Specht
275-285

Interactive Videoconference Supported Teaching in Undergraduate Nursing: A Case Study for ECG
Ufuk Celikkan, Fisun Senuzun, Dilek Sari and Yasar Guneri Sahin
286-294

Predicting Adolescent Deviant Behaviors through Data Mining Techniques
Yu-Chin Liu and Yung-Chieh Hsu
295-308

Effects of Thinking Style on Design Strategies: Using Bridge Construction Simulation Programs
Chuen-Tsai Sun, Dai-Yi Wang and Yu-Yeh Chang
309-320

A Systematic Understanding of Successful Web Searches in Information-based Tasks
Mingming Zhou
321-331

Predictive Effects of Online Peer Feedback Types on Performance Quality
Fu-Yun Yu and Chun-Ping Wu
332-341

Incorporating Usability Criteria into the Development of Animated Hierarchical Maps
Yu-Cheng Shih, Pei-Ren Huang and Sherry Y. Chen
342-355

Using Lean in the Flipped Classroom for At Risk Students
Shannon Flumerfelt and Greg Green
356-366

Using E-readers and Internet Resources to Support Comprehension
Sandra Wright, April Fugett and Francine Caputa
367-379

Design of Online Report Writing Based on Constructive and Cooperative Learning for a Course on Traditional General Physics Experiments
Hao-Chang Lo
380-391

Meeting the "Digital Natives": Understanding the Acceptance of Technology in Classrooms
Xiaoqing Gu, Yuankun Zhu and Xiaofeng Guo
392-402

Effects of Video Caption Modes on English Listening Comprehension and Vocabulary Acquisition Using Handheld Devices
Ching-Kun Hsu, Gwo-Jen Hwang, Yu-Tzu Chang and Chih-Kai Chang
403-414
Guest Editorial: Innovative Technologies for the Seamless Integration of Formal and Informal Learning

Patrizia Ghislandi¹, Maria Grazia Ierardi², Tommaso Leo³ and Luca Spalazzi³
¹Dipartimento Scienze Cognizione e Formazione, Università degli Studi di Trento, Trento, Italy // ²Ist. Matematica Applicata e Tecnologie Informatiche, Consiglio Nazionale delle Ricerche, Genova, Italy // ³Dipartimento di Ingegneria dell’ Informazione, Università Politecnica delle Marche, Ancona, Italy // patrizia.ghislandi@unitn.it // marygz@ge.imati.cnr.it // tommaso.leo@univpm.it // l.spalazzi@univpm.it

As multiple devices with advanced computing and networking capacities are now prevailing in our lives, teachers and researchers are seeking for new types of learning methods trying to integrate formal, informal and non-formal learning. These approaches provide the opportunity to expose everybody to learning occasions in a number of environments - tangible or virtual - where we can take advantage of these new technologies.

Looking at the wide set of new learning opportunities from the point of view of the professionals, the speed of change and the increased demand of flexibility in being able to build knowledge and to adapt themselves to new forms of collaboration and work, the capability of successfully exploiting the increasing variety and accessibility of learning occasions and resources, becomes a crucial issue.

Along with this, there is the ever growing interest in the use of social media and applications in order to realise bottom-up, participative alternatives to the traditional teaching/learning approach.

These are the promises of the use of innovative technologies for innovating learning/teaching approaches. But every promise involves some kind of risk. Mixing virtual and real environments could hamper the creativity of the children. The excessive availability of web resources could engender the syndrome of “lost in cyberspace” in the learners, young and adults. To introduce a new technology in a formal setting could result without sense for the target users if social aspects are not properly considered.

All this is the basis of the growing attention for rethinking the current view of e-learning, as far as the educational models and the corresponding technical solutions are concerned. This special issue focuses on surveying some current research activities in such a multifaceted scenario.

We are sad for missing before its launch our colleague Paola Forcheri, unfortunately killed by a cruel illness, who proposed the theme of the issue.

24 papers were submitted to this Special Issue, each characterised by a specific mix of pedagogical approaches, technologies taken into consideration, learning/teaching objectives and by the personal views of the Authors about terms like integration, innovation, formal and informal. It is interesting to note the distribution of the papers per Country: we had 3 papers from Malaysia, 1 from Turkey, 2 from Nigeria, 2 from Taiwan, 1 from Greece, 2 from Brazil, 2 from USA, 1 from United Arab Emirates, 2 from Algeria, 2 from Kuwait, 1 from Jordan and 6 from Italy.

We would only underline that the interest for the topic seems to be diffused all around the world.

We have to express our gratitude to the reviewers who carried out a great work (great both for the amount and for the quality) in doing their twofold double blind review, that resulted in 10 papers accepted.

We attempted to clusterize the 10 papers that actually compose this special issue according to different keys (technologies, objectives, pedagogical issue, and case studies vs. design or vision papers) and finally we decided to group them according to the target users, because of the null subjectivity of this classification.

The reader will find first the 5 papers addressing technologies for schools (Pereira et al., Di Blas and Paolini, Nury Kara et al., Ardito et al., Hayashi and Baranauskas), then the 3 papers considering young people (Al Nashad and Gunn, Then-Chi Huang and Chia-Chen Chen, Dettori and Torsani) and finally the 2 papers dealing with adult users (Manganello et al., Chaoui and Laskri)
In the first paper “Social Software and Educational Technology: Informal, Formal and Technical Values”, Roberto Pereira, Maria Cecilia C. Baranauskas e Sergio Roberto P. Da Silva focus on people’s values issue to inform the design of social software. The value issue is suggested to enforce the usual technical issue like privacy, reputation, and identity. Social software is meant as a technology usable to bridge the gap between informal and formal learning, i.e. to promote social interaction in learning contexts. Grounding their arguments on Organizational Semiotics and the Building Blocks of Culture, they suggest considering 28 values within the Value Pie.

In the second paper “Beyond the School’s Boundaries: PoliCultura, a Large-Scale Digital Storytelling Initiative”, Nicoletta Di Blas and Paolo Paolini discuss the effect of technologies in supporting students and teachers to go beyond the school boundaries, both in space and time, with respect to their mutual interaction and to the access to the external world. They consider the case of PoliCultura, a large-scale (20,000 users) digital storytelling initiative at the borders between formal and informal education, in which students and teachers collaboratively create a multimedia story. In order to accomplish this task, they interview experts, visit local institutions, involve their families and the community at large, cooperate through social media with remote peers, working at school as well as from home. Their analysis is aimed to provide evidence that, in doing so, students not only get engaged but they achieve substantial educational benefits.

The third paper “Investigating the Activities of Children toward a Smart Storytelling Toy”, by Nuri Kara, Cansu Cigdem Aydin, and Kursat Cagiltay, presents StoryTech, a smart storytelling toy that features a virtual space, including computer-based graphics and characters, and a real space, where the pupils have plush toys, background cards, and a communication interface. The paper also provides information on the attributes, design, and development process of StoryTech. When children put real objects on the receiver panel, the computer program shows proper backgrounds and characters on the screen and encourages children to produce their own stories. The results of an empirical research involving 90 children are given to indicate that StoryTech contributed to narrative activities of children and had a positive impact on their creativity.

By the fourth paper, “Integrating traditional learning and games on large displays: an experimental study”, by Carmelo Ardito, Rosa Lanzilotti, Maria Francesca Costabile and Giuseppe Desolda, we are driven to analyse the effects of a new educational format, inspired by the Discovery Learning technique, which integrates educational games, designed to be played on large multi-touch displays, with other types of formal and informal learning. The format comprehends 1) attending the lesson(s) in the classroom (symbolic representation), 2) acting in a real context (active representation), and 3) interacting with the multi-touch display to manipulate images (iconic representation). Six classes of a primary school were involved in the study. The results showed that the proposed educational format was effective, in particular with respect to the inclusion. Games on these novel multi-touch systems engage pupils very much, stimulate their collaboration and help consolidating knowledge.

In the fifth paper “Affectibility in Educational Technologies: a Socio-Technical Perspective for Design”, by Elaine C. S. Hayashi and Maria Cecilia C. Baranauskas, a large-scale experience of introduction of the XO educational laptop at an elementary public school in the city of Campinas, in São Paulo, Brazil, involving more than 500 people is presented. The research was conducted to address questions such as: How could a new technology be introduced in schools in a way that it makes sense to the users? Could it contribute to more integrated learning scenarios? Particular attention was given to new perspectives on understanding learning practices mediated by technology and to the role of the concept of ‘Affectibility’ in the design of educational systems.

Data were collected from participatory Workshops, informal interviews and pictographic questionnaire. The results suggest that it is possible to combine school’s formal and informal practices into meaningful learning. The activities helped teachers and students realize that the use of technology can be recreated, influenced by their own feelings, values and culture.

The following three papers refer to high school and university students.

“Lecture Capture in Engineering Classes: Bridging Gaps and Enhancing Learning” by Hasan Al Nashash and Cindy Gunn explores the use of lecture capture and webcasting in Engineering classes to promote formal and informal learning of the students by providing them with the opportunity to enhance autonomously their understanding of the course content. The study was conducted from the point of view of the students. They were asked to provide feedback on what they perceive the benefits and the drawbacks of lecture capture to be. It appears that the students consider lecture capture an effective tool to help them succeed in the course. The videos available at every time
allowed students to bridge the gap between what they have understood in the formal class setting and what they are able to better understand after reviewing the videos in a relaxed environment. In addition, most of the students indicated that the availability of the videos did not encourage them to skip or miss any classes. The main drawback was associated with technical difficulties which resulted in some wasted time.

The paper “Animating Civic Education: Developing a Knowledge Navigation System using Blogging and Topic Map Technology”, by Tien-Chi Huang, Chia-Chen Chen, presents an attempt to integrate educational blog articles and to use the wealth of information inherent in the development of a blogging system with knowledge navigation demand for both formal and informal learning in civic education in high schools. In the system, Topic Maps (TM) technology was implemented to represent informal learning content in a formal curriculum structure. The study proposes the use of Semantic Web to assist in Knowledge navigation as a seamless model for teachers and system developers to represent informal learning content that adheres to the formal knowledge structure. Analysis on the field indicates that students approved the effectiveness of combining the blogging system and individual Topic Maps (iTM) function. In other words, the system derived from the design framework helps students learn citizenship courses in an innovative manner.

Giuliana Dettori and Simone Torsani with the paper “Enriching Formal Language Learning with an Informal Social Component”, provide an insight on the use of data mining and social bookmarking to build up an easy-to-use tool for enhancing formal language learning with an informal component. This is able to help the learners reach relevant Internet pages they can freely use to complement their learning activity. Thanks to this facility, each lesson is enriched, at run time, with a number of links automatically retrieved from social bookmarking sites. The learners also have at disposal a micro-evaluation system allowing them to share within the environment their like or dislike of the visited web sites, giving their fellow learners indirect guidance for fruitful web exploitation. The analysis of the retrieved bookmarks for different kinds of learning contents showed that suitable links were actually retrieved by this facility. Moreover, a pilot experimentation revealed that students actually feel inclined to make use of this facility and appreciate both the proposed links and the possibility to receive advice from their peers by means of the micro-evaluation system. This suggests that the added facility actually constitutes an opportunity of informal learning suitably connected with formal one.

The last two paper deal with an adult target population.

The paper “PKS: an Ontology-based Learning Construct for Lifelong Learners”, by Flavio Manganello, Carla Falsetti, Luca Spalazzi, and Tommaso Leo, addresses adult lifelong learners, i.e. persons interested in learning or compelled to learn during their working life but not able to, or not interested in participating in formal learning. These learners are considered motivated and self-aware enough to self-direct their learning, are presumed to be novices with respect to the needed knowledge and have a limited technological uptake. In the paper their main differences from regular learners, in particular, for the use of social media to improve learning skills, are outlined. Andragogy is assumed as the reference pedagogical model. A Service-Oriented Architecture, named Personal Knowledge Space (PKS), is proposed to support such lifelong learners in selecting, organizing, and retrieving information; in streamlining interaction processes among learners, services and resources and finally in empowering control and trust of personal relationships born during the learning processes. This way, formal and informal learning resources are collected and organized to build up significant knowledge. The PKS is mainly based on the exploitation of semantic tools (ontologies) and web services. Use cases describing the PKS architecture and one scenario of PKS use are presented.

The last paper in the issue is “Proposition and Organization of an Adaptive Learning Domain based on Fusion from the Web” by Mohammed Chaoui and Mohamed Tayeb Laskri. It addresses the objective of making easier for the teachers the recovery of web resources (informal resources) in order to build up adaptive learning resources to be offered on e-learning platforms for the purpose of formal learning. They proposed a Fusion of Web resources approach to this problem and one Artificial Intelligence based tool to organize resources to an Adaptive Learning Domain into e-learning platform. The focus was firstly on searching tools and filtering methods to extract the most relevant educational Web resources and structuring them to create courses, secondly on adaptation of extracted Web resources. The Authors outline a new process of fusion in creation and in adaptation to learner’s profiles. The process neither asks much time nor much effort from the authors to create courses. The process provides also direct way to the needed Web resources. Courses can be updated directly from the Web by reuse of the previously extracted and selected resources. The results of the tool evaluation give high performance in comparison with different methods of course's creation.
Social Software and Educational Technology: Informal, Formal and Technical Values

Roberto Pereira¹, M. Cecília C. Baranauskas¹ and Sergio Roberto P. da Silva²

¹Institute of Computing, University of Campinas (UNICAMP), São Paulo, Brazil // ²Department of Informatics, State University of Maringá (UEM), Paraná, Brazil // rpereira@ic.unicamp.br // cecilia@ic.unicamp.br // sergio.r.dasilva@gmail.com

ABSTRACT

Social software is a growing reality worldwide and several authors are discussing its use to promote social interaction in learning contexts. Although problems regarding privacy, reputation, and identity are commonly reported in social software, an explicit concern regarding peoples’ values is not a common practice in its design and adoption, in part, due to the lack of research in this subject. The issue of values becomes even more critical as social software crosses the boundaries of people’s cultures to pervade every aspect of their lives, from personal relationships to work, from play to education. In this paper we shed light on this scenario by presenting an informed discussion about values in the context of social software, as it may bridge the gap between informal and formal learning. An organization of 28 suggested values is presented in the Value Pie, as a way of informing the design of social software. Our discussion is grounded on Organizational Semiotics and the Building Blocks of Culture.

Keywords
Organizational Semiotics, Culture building blocks, Design, Analysis

Introduction

The Web 2.0 was a milestone in the development of rich and innovative web systems in terms of interactivity, enabling the emergence of the so-called social software (e.g., social networks, wikis, and social bookmarking). Social software is frequently cited as determinant of transformations that are changing the way people relate to digital technology (Pereira et al., 2010). Twitter®, Facebook® and Youtube® are examples of how information and communication technology (ICT) has pervaded every aspect of people’s personal and social life. This kind of system is used not only at home, but also in workplaces, public organizations, and schools for several purposes, via different devices and with far-reaching consequences.

Sellen et al. (2009) assert that we now live with technology, not just use it. What this means is that a broad set of factors that range from emotion, affect, sociability and human values, to issues of scalability, security and performance, have changed the interaction between people and computers. In fact, interactions have assumed new dimensions and cannot be addressed only as being task-oriented. The concepts we mentioned above, such as human values, motivation, pragmatics, emotion, and affect, that were traditionally left on the margin of approaches to computer systems development, need to be moved to the centre in order to develop systems aligned with the new demands of a society mediated by ICT (Harrison et al., 2007).

Authors such as Chatti et al. (2007), Dalsgaard (2006), Dron (2007), and Klamma et al. (2007), discuss the adoption and use of social software to promote social interaction for both informal learning and distance learning — traditionally centered on Learning Management Systems (LMS). Although these authors have different approaches, and focus on different aspects of the integration between informal and formal learning, they are congruent in suggesting social software as a technical solution for this issue. However, as social software remains a quite unexplored topic by the research community, the understanding of such a complex integration still demands further investigation.

The concept of social software, and the changes and challenges it brings, are being discussed in the literature and through informal discussions in forums and blogs since mid 2004. In Pereira et al. (2010), we presented a review and compilation of discussions indicating the need for a paradigm shift in the way we understand and design social software. As social software, and web applications in general, are available worldwide, we argue that the various elements should be understood as values bounded to cultural aspects of people, groups, organizations, and their environments, which are manifested in the informal, formal, and technical levels of information. As a challenge, we...
pointed out the need for studies, investigations, and theories to support understanding and placing values at the core of the analysis and design of social software.

According to Rokeach (1973), the value concept seems to be able to unify the apparently diverse interests of the sciences concerned with human behavior. Boyd (2007) asserts that social software is all about the new in web applications, but the new is more related to people’s behavior than to the technology itself. Social software introduces many complex issues that pervade every aspect of people’s lives, representing opportunities and challenges, benefits and drawbacks, democracy and exclusion. Therefore, taking into account values in social software design is among the most complex scenarios we are facing nowadays. If we consider the design of the systems previously cited, there is little concern for human values such as privacy, reputation, autonomy, among other cultural aspects (e.g., beliefs, behavioral patterns). An evidence of such negligence of the social aspects of these applications, users have been inadvertently serving as beta testers of applications as well as subjects of implicit behavioral experiments to identify the viability of a resource or product. Privacy policies and agreement terms are constantly changed and updated, many times without users’ awareness. Products which were not approved in their beta tests are removed/discontinued disregarding possible effects on users. Accessibility issues are usually neglected, making it difficult or even preventing the access of people that do not fit the myth of the “average user”.

Other evidences of implications related to values (or to the lack there of) in social software design have been widely reported by media, and can be easily perceived. Solove (2007) was one of the first authors to devote attention to the possible negative effects of social software on people’s privacy and reputation. However, as Sellen et al. (2009) highlight, human values, in all their diversity, should be seen according to the way they are supported, promoted, or inhibited by technologies. Specifically, in the context of social software, little is known about values. What these values look like, their roles, the way they can be promoted or inhibited, and the possible means to deal with them still demands investigation.

In this paper we shed light on these challenges by presenting a literature review on values in social software and educational technology, suggesting 28 elements that we approach as interactive values. Grounded on theories such as Organizational Semiotics (Liu, 2000), the Building Blocks of Culture (Hall, 1959), and Values Theory (Schwartz, 2005), we propose the Value Pie: an informed scheme that presents values as bounded to culture and manifested in the informal, formal, and technical layers of information systems.

The paper is organized as follows. First, we present a discussion on values in technology design. Second, we suggest, describe, and exemplify 28 values identified through a literature review on social software and educational technology. Third, we propose the Value Pie and exemplify our discussion in the context of informal and formal learning. Finally, we present our final remarks.

**Values in Design**

Friedman et al. (2006) define values as something that is important to a person or group of people. In the Values Theory, Schwartz (2005) defines values as desirable, trans-situational goals that vary in importance and that serve as principles that guide people’s lives. In this theory, values are understood as beliefs tinged with emotions, as motivational constructs that transcend specific situations and actions, serving as standards or criteria to guide the selection of actions, policies, people, and events.

For Knobel and Bowker (2011), values often play out in information technologies as disaster needing management, because conversations and analysis of the values in technologies generally occur after their design and launch. Consequently, most users are faced with undecipherable and sometimes strange decisions previously made on their behalf, often not to their benefit. Friedman (1996) argues that the cost to disseminate a technology is insignificant when compared to the cost of developing it, so that the values embedded in its implementations are deep, systematic, and easily disseminated. According to the author, although the negligence of moral values in any organization is disturbing, it is particularly damaging in the design of computer technology, because, unlike people with whom we can disagree and negotiate values and their meanings, we can hardly do the same with technology. Therefore, when designing computer technology it is necessary to see human values from an ethical point of view.
In the context of educational technology, researchers such as Chatti et al. (2007), Chen et al. (2009), Dalsgaard (2006), and Roberts and McInerney (2007), are convinced that systems such as LMS are not capable of supporting learners in an interactive and effective learning process. The authors highlight the inability of educational technologies to promote a continuous social interaction, due to their focus on content instead of people, and the lack of resources to motivate users’ participation and interaction outside the limits of a formal course or institution. On the one hand, Dalsgaard (2006) points out the efficiency of such systems regarding administrative issues, its organization in a top-down format, and the management of courses and their activities. On the other hand, Chatti et al. (2007) criticize the one-size-fits-all model adopted by LMS. These authors understand learning as a social process and argue that educational technologies usually treat learning just as “courses delivery”, “learning objects”, and “learning resources”, carrying to the virtual environment the procedures, structure, and activities already existent in traditional learning models in institutions. The cited authors are unanimous in suggesting social software as an interesting alternative to promote learning in informal contexts, as well as to foster social interaction and users’ motivation in combination with educational technologies.

Although the previously cited works have pointed out the viability of using social software to promote the integration of informal and formal learning, they are usually concerned with technical issues and/or formal procedures and practices of teaching/learning, not approaching informal learning directly. There is a lack of research that takes into account the impacts of this integration on the values of the stakeholders involved. Understanding the way values are intertwined with the informal, formal, and technical aspects of learning is key to produce educational technologies capable of integrating these aspects.

Some works have explicitly focused on values in technology design. Cockton (2005) proposes a framework to support what he named a Value-Centred Design, which suggests some activities and artifacts to support designers in the understanding of technology design as a process of delivering value. Adopting a different perspective, during the last decade Friedman (1996) has been working on an approach she named Value-Sensitive Design, which is intended to support the concern with human values in the design of computer systems, especially the ethical ones. Sellen et al. (2007) assert that presently the biggest challenge in the design of technology is the explicit consideration of values. Their assertion reinforces the choir of Miller et al. (2007), Knobel and Bowker (2011) and Bannon (2011) who argue for studies, methods, artifacts, and examples to support designers in understanding values in technology design.

A Survey on Values in Social Software and Educational Technology Design

Values are bound to culture (Hall, 1959; Schwartz, 2005). They are intertwined with each other and vary in meaning, importance, and priority according to the culture being analyzed and across time and space. In the context of computer systems, depending on the way the system is designed, it may afford behaviors that are intrinsically related to individuals and the complex cultural context in which they are using it.

Hall (1959) understands culture as a form of communication and, in an attempt to formalize its basic constructs, he proposed 10 areas, or culture building blocks, he named Primary Messages Systems (PMS): Interaction, Association, Learning, Play, Protection, Exploitation, Temporality, Territoriality, Subsistence and Bisexuality — in Liu (2000) approached as “Classification”. Each area is biologically rooted, and any culture can be seen as an evolution of human behaviors and interactions mapped by a combination of them. For him, all cultures develop values with regard to these 10 areas. For instance, values in “Association” refer to the way people organize and structure themselves in society; the groups they form, the kind of relationship they develop and maintain, etc. “Family” is a manifestation of the aspect of “Association” in a given culture, and its role, structure, and relative importance in society can be understood as values developed in/by that culture. Values in “Classification” refer to preferred style of dressing, jobs, sports, and so on, of men and women. In the same way, values in “Learning” may be related to valued abilities, knowledge and professions; the relative importance of experience, expertise, meritocracy, and others.

In Pereira et al. (2010), we presented analyses of existing social software as well as a review and organization of some discussions about the subject. In that work, we identified and suggested 13 elements we called the social software building blocks, drawing attention to the need for leaving a technically-centered perspective in favor of one that encompasses and articulates informal, formal, and technical aspects involved in social software.
In this paper, we revisit and expand our literature review on social software and educational technology in order to identify, discuss, and exemplify what we are calling informal, formal, and technical values. For the literature review, we selected three journals and three conferences according to their tradition and importance in the areas of Human-Computer Interaction (HCI) and Educational Technology (due to the growing discussion about the design of technology to support teaching and learning that promote social interaction). The International Journal of Educational Technology & Society, the Computers & Education Journal and the British Journal of Educational Technology were selected based on their tradition, scope, and societal reach. The International Conference on Human-Computer Interaction (HCII), the Conference on Human-Computer Interaction (IFIP TC13-INTERACT), and the Conference on Human Factors in Computing Systems (ACM CHI) were selected in order to consider three of the most important and comprehensive international conferences in the field of HCI.

Initially, the investigation considered all the scientific papers published from 2005 to 2010 at the conferences, and the scientific papers published in the same period in the journals, based on the following keywords: web 2.0, social web, social software, social network, values and life-long learning. Additionally, for the International Journal of Educational Technology & Society, the search was expanded to encompass papers published from 2000 to 2011. Papers published in other journals and conferences which were referenced by the selected papers were also considered — see Table 1. This initial survey resulted in a total of 2,300 papers. In the first stage, 161 papers were selected based on title and abstract. In the second stage, a detailed analysis of the pre-selected papers resulted in the identification of 65 papers relevant to our interest in values, elements, success factors, and guidelines for designing social software and educational technology that promotes social interaction.

<table>
<thead>
<tr>
<th>Name</th>
<th>Period</th>
<th>Papers Selected</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFIP-INTERACT</td>
<td>Conference on Human-Computer Interaction</td>
<td>2005, 2007, 2009</td>
<td>4 1</td>
</tr>
<tr>
<td>C&amp;E</td>
<td>Computers &amp; Education</td>
<td>2005 - 2010</td>
<td>9 3</td>
</tr>
<tr>
<td>BJET</td>
<td>British Journal of Educational Technology</td>
<td>2005 - 2010</td>
<td>6 2</td>
</tr>
<tr>
<td>Other journals and conferences</td>
<td></td>
<td></td>
<td>41 37</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>161</strong></td>
<td><strong>65</strong></td>
</tr>
</tbody>
</table>

Regarding the literature on learning technologies, Klamma et al. (2007) discuss the necessary characteristics of collaborative adaptive learning platforms; McLoughlin & Lee (2007) present 12 examples of pedagogical innovations using social software in 4 countries; Dron (2007) proposes 10 principles to the design of educational social software; Roberts & McNerney (2007) point out seven problems of online group learning and suggest their solutions. These authors consider social software within formal learning contexts where teachers and students have an explicit relationship with an educational institution. They usually focus on technical aspects of social software and formal aspects of learning. The nature of social software is informal. Thus, it may support informal learning through social interaction taking advantage of its ability to deal with users’ experience attributes, such as pleasure, motivation, and creative involvement.

Grounded in the relevant papers and in our previous work, we identified 28 elements that represent critical aspects and could be seen as values in systems intended to promote social interaction. Table 2 presents these values, their description, including a formal definition of them by Britannica (2011) — in italic, and the most meaningful references whose discussions allowed their identification.

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>“Providing access; capable of being reached; also being within reach; easy to communicate or deal with”</td>
<td>(Hernández-Ramos, 2006)</td>
</tr>
<tr>
<td></td>
<td>Capability of supporting, satisfactorily, a heterogeneous set of users with distinct skills, preferences, needs, perceptual and learning disabilities, and motor and cognitive limitations. Example: the system is fully readable by a</td>
<td>(Almeida et al., 2009)</td>
</tr>
<tr>
<td>Screen-reader application.</td>
<td>“Capable of being or becoming adapted” Possibility of modifying a system according to its context of use; flexibility of being adapted to different contexts; situations of use that have changed or are unexpected. Example: users can create contact lists and configure the information they want to share with each other. (Dron, 2007) (Neris et al., 2007)</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Adaptability</td>
<td>“External show; outward aspect; outward indication; a sense impression or aspect of a thing” Characteristics related to attractiveness, beauty, care with the image, and the way things are displayed and presented. Example: standardized interfaces with significant and well designed graphic elements. (Lazar &amp; Preece, 2003) (Norman, 2008)</td>
<td></td>
</tr>
<tr>
<td>Aesthetic, Appearance</td>
<td>“The quality or state of being self-governing; especially the right of self-government” Users’ ability to decide, plan, and act in a way they believe help them reach their goals. Ability to control the technology and use it to their advantage. Example: features that make the system self-explanatory for novice users. (Bannon, 2011) (Friedman, 1996)</td>
<td></td>
</tr>
<tr>
<td>Autonomy</td>
<td>“The quality or state of being available”; Available: “present or ready for immediate use” Refers to the capability of the system, feature, or functionality of being available for use at any time and without interruption. Example: the system is available 24/7. (Isaias et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>Availability</td>
<td>“Watchful, wary; having or showing realization, perception, or knowledge” Individual and/or collective perceptions about who is available in the system; who is doing what; what is happening and what happened, etc. Example: users are notified about the news existent since their last logon. (Chatti et al., 2007) (Glahn et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>Awareness</td>
<td>“To work jointly with others or together especially in an intellectual endeavor”. Possibility for cooperating; working together on the same object. Example: users jointly create, edit, and evaluate an article in a Wiki. (Knobel &amp; Bowker, 2011) (Vavoula et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td>“Oral exchange of sentiments, observations, opinions, or ideas” Possibility of two or more users establishing direct communication (synchronous and/or asynchronous). Example: comments, chat. (Tsai et al., 2008) (Vavoula &amp; Sharples, 2009)</td>
<td></td>
</tr>
<tr>
<td>Conversation</td>
<td>“A conscious mental reaction (as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body” Feelings, people’s sensations such as welfare, pleasure, fun, engagement, boredom, disappointment, and other aspects related to the user’s experience. Example: users’ fear of suffering discrimination because of information published on their profile; features that allow users to express their affective state. (Brandtzaeg &amp; Heim, 2009) (Norman, 2008) (Almeida et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>Emotion and Affection</td>
<td>“A number of individuals assembled together or having some uniting relationship” A set of individuals with characteristics, situations, interests, or purposes in common. Example: a group of people interested in the use of social software in education. (Roberts and McInnerny, 2007) (Tsai et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>Groups</td>
<td>“The distinguishing character or personality of an individual” The “self” of individuals; the expression of elements of a person’s personality and individuality (who the person is over space and time). Example: the representation of a person’s profile, their activities, personal information, etc. (Boyd &amp; Elisson, 2007) (Knorr-Cetina, 1997) (Pereira et al., 2010)</td>
<td></td>
</tr>
<tr>
<td>Identity</td>
<td>“Consent to surgery by a patient or to participation in a medical experiment by a subject after achieving an understanding of what is involved” Users’ awareness about the possible impacts of their actions. Refers to informing and garnering people’s agreement about what is produced from their interaction with the system and with other users. Example: a user agrees to make his/her profile public even after the system alerting about the risks of (Friedman et al., 2006) (Miller et al., 2007)</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Definitions and Examples</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Meta-communication</strong></td>
<td>Meta: “occurring later than or in succession to; situated behind or beyond; later or more highly organized or specialized form of”. Communication: “an act or instance of transmitting; a process by which information is exchanged between individuals through a common system of symbols, signs, or behavior” Features that make the system self-explanatory; that allow the user to understand how the system works, the reason it was designed in that way, what can be done through the available resources/features, and what are the possible impacts of using them. Example: the system has explanations and offers tips recorded in video and sign language that guide the user regarding privacy settings. (Hayashi &amp; Baranauskas, 2010)</td>
<td></td>
</tr>
<tr>
<td><strong>Norms, rules and policies</strong></td>
<td>Norm: “a principle of right action binding upon the members of a group and serving to guide, control, or regulate proper and acceptable behavior” Formal aspects that govern, regulate, and determine how individuals behave, think, make judgments, and perceive the world. Example: the system’s terms and conditions of use. (Neris et al., 2007) (Lazar &amp; Preece, 2003)</td>
<td></td>
</tr>
<tr>
<td><strong>Object</strong></td>
<td>“Something mental or physical toward which thought, feeling, or action is directed” Artifacts around which social interactions occur (e.g., the talks, the focus, the collaboration, etc.). Example: videos on YouTube, short messages on Twitter, comments, sharing on Facebook. (Knorr-Cetina, 1997) (Pereira et al., 2010)</td>
<td></td>
</tr>
<tr>
<td><strong>Ownership and property</strong></td>
<td>“The state, relation, or fact of being an owner” The right to possess an object or information, and the right over the actions that can be executed over/with/through this object. Example: a user creates a document, changes it, and shares and transfers its ownership to another user. (Friedman, 1996) (Sellen et al., 2009)</td>
<td></td>
</tr>
<tr>
<td><strong>Portability</strong></td>
<td>“The quality or state of being portable”. Portable: “usable on many computers with little or no modification” Possibility of using the system, its features and functionalities, through different devices and on different platforms. Ex: access through mobile phone, compatibility with different browsers. (Isaias et al., 2009) (Sellen et al., 2009)</td>
<td></td>
</tr>
<tr>
<td><strong>Presence</strong></td>
<td>“The fact or condition of being present” Refers to a person being (or not) in a certain place at a certain time. Example: the user is online in the system. (McLoughlin &amp; Lee, 2007) (Pereira et al., 2010)</td>
<td></td>
</tr>
<tr>
<td><strong>Privacy</strong></td>
<td>“The quality or state of being apart from company or observation; the right to freedom from unauthorized intrusion” A claim, entitlement, or right of users to determine what information about them will be available and who has access to that information. Example: the system allows users to show his/her list of friends only to people who are already part of it. (Cotler &amp; Rizzo, 2010) (Glahn et al., 2009) (Karat et al., 2008)</td>
<td></td>
</tr>
<tr>
<td><strong>Reciprocity</strong></td>
<td>“The quality or state of being reciprocal; mutual dependence, action, or influence; a mutual exchange of privileges” Feeling of mutual benefit; reward for performing a task or for employing some effort to achieve a goal. Example: users that provide high quality contents appear in the “top contributors’ users”. (Chen et al., 2009) (Glahn et al., 2009) (Klamma et al., 2007)</td>
<td></td>
</tr>
<tr>
<td><strong>Relationship</strong></td>
<td>“The state of being related or interrelated; a specific instance or type of kinship; a state of affairs existing between those having relations or dealings” Some kind of link or social tie between two or more individuals. Example: followers, fans, friends. (Boyd &amp; Elisson, 2007) (Karat et al., 2008)</td>
<td></td>
</tr>
<tr>
<td><strong>Reputation</strong></td>
<td>“Overall quality or character as seen or judged by people in general; recognition by other people of some characteristic or ability” The perception or collective opinion about an individual constructed by others. Example: the user is an expert in a specific area; the user is known for sending spam and undesirable content. (Bannon, 2011) (Solove, 2007)</td>
<td></td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>“Capable of being scaled; capable of being easily expanded or upgraded on” (Boyd, 2007)</td>
<td></td>
</tr>
</tbody>
</table>
The list presented in Table 2 adds to the previously cited findings by considering not only technical and formal issues involved in the context of social software and educational technologies, but also the informal ones. It also includes values particularly relevant to educational technologies. For instance, “meta-communication” has shown to be effective for promoting users’ autonomy (Hayashi & Baranauskas, 2010). “Informed consent” seems to be important in dealing with conflicts between privacy and visibility favoring security (Friedman et al., 2006). Social conventions, rules, procedures, laws etc., are different “norms” that govern how society works and must be explicitly considered. “Accessibility” issues must be a main concern in the design and adoption of any technology. “Emotional and affective” aspects involved in educational technology development, adoption, and use also need to receive attention. The concern with these values seems to be even more critical when we consider informal learning, where there is no formal assistance and guidance; where users need to be autonomous, aware of their possibilities, and the consequences of their actions; where users need to be motivated, confident, and rewarded for their efforts using the technology and interacting with each other. If these values are neglected, we run the risk of importing to educational technologies the problems widely reported in social software (e.g., privacy, reputation, exclusion) without fully taking advantage of its use to encourage the integration between formal and informal learning.

We must highlight, however, that the 28 values do not constitute a definitive or exhaustive list. Indeed, as Friedman et al. (2006) argue, perhaps no list could be, at least in such a broad and complex area. For example, more abstract values such as solidarity, well-being, involvement, motivation, satisfaction, and aspects of user’s experience are represented by the value “Emotion and affection”. Some values may be decomposed (e.g., “collaboration” encompasses “cooperation”, “security” encompasses “safety”) or are transversal to the others (e.g., “norms”). Our main concern when creating this list was to find a balance between making it as comprehensive and diverse as possible without having it be overly complex or detailed.

An Organized Scheme for Values

We previously suggested the importance of taking values into account when discussing informal and formal learning. According to Hall (1959), humans operate at three different levels: informal, formal, and technical. In the learning context, for example, people may learn from observing other people and imitating them (informal); from other’s
explicit feedback, suggestions, and instructions (formal); or from books, guidelines, and other materials that explain and justify things in a coherently outlined form (technical). A given culture may emphasize technical learning while another may be heavily informal. As Hall asserts, we can identify all the three levels in almost any learning situation, but one will always be emphasized.

It is possible to perceive the three levels in action when the adoption of social software to foster educational practices is being considered. There are several informal issues at play, mainly the emotional and affective ones such as students’ motivations and teachers’ openness to change. There are also formal issues that must be understood and followed, such as the laws, the teaching program, and the students’ minimum age. And there are technical issues, which range from choosing the right social software (that respects the formal issues and is in conformity with the informal ones) to the physical structure (space, internet access, network security).

These three levels are also related to the values. For instance, “autonomy” and “identity” are clearly informal issues, while “norms, rules and policies” are clearly formal, and “scalability” and “portability” clearly technical. In this sense, the Organizational Semiotics theory (Liu, 2000) proposes the Semiotic Onion to explain how these levels exist in the context of organizations and information systems. The informal represents the organizational culture, customs, and values that are reflected as beliefs, habits, and individual behavior patterns of its members. The formal corresponds to aspects that are well established and accepted, becoming social conventions, norms, or laws. In this level, rules and procedures are created to replace meanings and intentions. Finally, the technical, situated at the core of the onion, represents aspects that are so formalized that they can be technically approached and supported. In a social perspective to the design of computing systems, Baranauskas (2009) and Baranauskas & Bonacin (2008) draw attention to the need for considering the three levels in an integrated and interrelated way in order to produce systems that make sense to their users, not causing negative impacts on them and the environment in which they will be used.

Aiming at situating the values at the three levels in which humans operate and in accordance to Hall’s (1959) areas of culture, we draw on the Semiotic Onion (Liu, 2000) and Schwartz’s Circular Model of Values (2005) to conceive the Value Pie — see Figure 1. The Value Pie is divided into 10 slices, each one corresponding to an area of culture, and is composed by three layers: the external layer refers to the informal level, the middle layer refers to the formal level, and the internal layer refers to the technical level. Values placed at the informal level usually have a personal or ethical nature; values situated at the formal level are collective or social values where there is a social rule or system of norms; and values placed in the technical level can be understood as quality attributes or special features of technology.

These values have an interactive nature and each level must not be approached in isolation. For instance, “autonomy” (informal level) may be restricted/promoted depending on the existing “norms and rules” (formal) and on “accessibility” issues (technical). Moreover, Hall (1959) explains that although one level always dominates and although we deal with them separately, the levels are simultaneously present in everything. For instance, based on Schwartz’s Circular Model of Values (2005) one may relate “privacy” to aspects of safety, harmony, and stability of the self, which are usually treated as informal concepts. Everyone has his/her own informal understanding of what privacy is and what it means. However, there are social protocols, conventions, rules, and laws that are formally established to define the meaning, limits, and guarantees of an individual’s privacy. Finally, there are also some facets of privacy that are so formally accepted that they can be technically supported, such as a curtain to cover a window or a feature for restricting the visibility of personal data in a social network website.

On the other hand, values placed in a same slice tend to have a natural congruence because they are developed in the same area of culture. For instance, all values developed in “Association” have some individual-individual/individual-object (Knorr-Cetina, 1997) association in their nature. “Group” is an association. “Relationship” is a kind of association. “Conversation” usually occurs when there are individuals associated in some way. “Trust” is built on and may reinforce association while the lack of trust may destroy it. Mapping it to a learning context, we can say that the existence of an explicit relationship between two or more individuals (e.g., personal contacts) may favor the emergence of a group (e.g., users interested in games), which may promote conversation among the participants (e.g., forum, chats), and which, in turn, may reinforce trust among the users (e.g., sharing ideas).

The Value Pie is not intended to be a classification scheme in which the elements are assigned to one and only one class within a system of mutually exclusive and non-overlapping classes (Jacob, 2004). Values may be developed at the intersection of multiple areas. For instance, “privacy” is developed at the intersection of “Protection-
Territoriality”. It appears in the “Protection” area because while the aspect of space changes (physical, personal) the aspect of protecting the space remains. Other values, such as “identity” and “norms”, are transversal to the 10 areas. For instance, one may see “identity” as the sum of an individual’s aspects, values, and behavioral patterns, related to the 10 areas (e.g., his/her position in a social group, his/her job, preferences in playing and learning). The value of “norms”, on the other hand, is present in the formal aspect of all the other areas, e.g., learning institutions and their rules (“Learning”), geographical limits and registry of property (“Territoriality”), time zone (“Temporality”), age defining adulthood (“Classification”), etc. The values of “identity” and “norms” appear in the “Interaction” area because it is also transversal to the other areas. For Hall (1959), interaction is at the center of culture and the other areas grow from it: interacting with the environment is to be alive, failing to do so is to be dead; everything people do involves interaction with someone/something else. In this sense, the Value Pie aims at organizing values according to their dominant PMS, and the unfilled spaces may indicate opportunities for reflection and for challenging designers and analysts.

![Figure 1. The Value Pie](image)

**Conclusion**

Usually related to informal contexts, social software has been regarded as bringing both opportunities and challenges to the academy as well as to governmental institutions and private organizations. Researchers of educational technologies were among the first to consider its use in formal settings, evidencing the need for making explicit the values involved in such a complex context.

Although recognized as important, there are few initiatives relating values to technology. In social software, there is even a lack of theoretically grounded approaches for investigating it. In this paper, we presented a survey of social software and educational technology in order to identify elements that should be considered by designers and practitioners when designing or adopting these systems for different usage contexts. As a result, 28 elements were identified and approached as values. Grounded in theories from different areas, we conceived the Value Pie: an informed organization scheme that presents values as bound to culture and manifested in the informal, formal, and technical layers of information systems. The list of values and the Value Pie are a first step in the direction of a value oriented and culturally informed approach to the design of technology intended to promote learning through social interaction. By drawing attention to the diversity of values and their interactive nature, they may be helpful in guiding designers, analysts, and practitioners to consider values when designing their systems.
Acknowledgements

This research is partially funded by CNPq through the EcoWeb Project (#560044/2010-0) and FAPESP (#2009/11888-7).

References


Beyond the School’s Boundaries: PoliCultura, a Large-Scale Digital Storytelling Initiative

Nicoletta Di Blas* and Paolo Paolini

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

*corresponding author

ABSTRACT

Technologies are changing the way we teach and learn in many respects. A relevant and not yet fully explored aspect is that they can support, even entice, students and teachers to go beyond the school boundaries, in spatial and temporal terms. Teachers and learners can keep in touch and work together, when they are not at school; they can access “the world” via Internet; peer to peer remote cooperation is possible; multimedia possibilities provide an incentive to explore the territory, the features of which can be documented in an effective way; digital content can be accessed, created, refined at any moment, at school and from home. This paper discusses this issue at the light of PoliCultura, a large-scale (20,000 users) digital storytelling initiative at the borders between formal and informal education, in which students and teachers collaboratively create a multimedia story. In order to accomplish this task, they interview experts, visit local institutions, involve their families and the community at large, cooperate through social media with remote peers, working at school as well as from home: in other words, they go “beyond the school’s boundaries”. In doing so, they not only get engaged but they achieve substantial educational benefits.

Keywords

Digital Storytelling, Educational Benefits, Multimedia, Multichannel, Collaboration, Authoring Tool

Introduction

Information and Communication Technologies (ICT) are dramatically changing the way we teach and learn in many respects. They are helping to diversify the way lessons are given, thanks to multimedia interactive content; they are improving the quality of contact with students and families; they are supporting a number of evaluation and assessment processes. One interesting and not fully explored yet outcome is that ICT can actually support, even entice, students and teachers to go beyond the school boundaries, both in spatial and in temporal terms. In a connected world, people can learn “anytime, anywhere” at “any path, any pace”, as the motto of the Florida Virtual School (FVS) goes. The Florida Virtual School is a striking example of innovation in this sense: it is an accredited, public, online e-learning school serving students in grades K-12. In theory, a student may take the whole program online without ever stepping into a classroom; in practice, only 1% of the FVS students are exclusively virtual; it rather happens that students join specific courses to supplement their studies at school, especially when they are underperforming (http://www.flvs.net).

This paper introduces an ICT-based initiative that uses technology to “go beyond the school’s boundaries”: PoliCultura, a large-scale deployment of collaborative digital storytelling at school, with 20,000 students (at June 2012), aged between 5 and 18, involved so far (www.policultura.it). PoliCultura pushes students and teachers to gather materials from sources other than the textbooks, to interact with their territory (institutions, authorities, experts…), to involve families (parents, grandparents), to communicate with remote peers through social media during and beyond school hours, in the end conjuring up highly innovative multimedia and multichannel digital stories (figure 1) that are shared through an international portal (figure 6). PoliCultura mingles formal and informal educational aspects: the whole activity is under the teachers’ supervision and framed within the curriculum, but at the same time students are pushed to organize themselves and to interact with the “external world”. The process of story-creation fosters the achievement of substantial educational benefits, many of which related to this “opening”, which is facilitated and supported by technology.

The paper is organized as follows: after a section on the “state of the art” of digital storytelling for youngsters (section 2), we plunge into the PoliCultura initiative’s description (section 3), with a specific focus on selected examples where the breaking of the school’s boundaries is quite evident (section 4). Some data on the evaluation are reported in section 5, but the reader is reminded that a full account of the educational impact is to be found elsewhere (Di Blas, Paolini, 2012). Eventually in section 6 we draw the conclusions and highlight our future steps of research.
State of the Art on Digital Storytelling

Digital storytelling (plus or minus the adjective “interactive”) is quite a huge field encompassing different phenomena and very different approaches (Iurgel et alii, 2009).

In this paper we focus on the educational aspects of digital storytelling in formal education, with students as authors rather than consumers. By digital storytelling we mean “the modern expression of the ancient art of storytelling” where “digital stories derive their power by weaving images, music, narrative and voice together, thereby giving deep dimension and vivid color to characters, situations, experiences, and insight” (definition by the the Digital Storytelling Association; http://electronicportfolios.com/digistory/). The Association for Progressive Communication further clarifies the concept: “Digital stories are stories produced, stored and disseminated using digital media. The focus is on the story tellers’ control over the medium, choice of words (narration), pictures and music so that the process is as powerful for the story teller as the end product is to the listener.” (http://www.apc.org/en/node/10567).

Authoring tools for digital storytelling have been mainly developed for very young children, being the educational value of authoring a story strongly backed by those pedagogical theories that consider learning as knowledge building rather than knowledge transmission (Feher, 2008; Jonassen, Land 2000). A number of tools have been developed to support this activity, both in the academic and commercial arena. Still, commercial products tend to see users more as listeners than authors; if they are considered as authors, then they are generally provided with readymade characters with which they can play role-games at most. These products are often CD-rom based, they impose strong limitations to creativity and almost never allow cooperation or sharing of the stories with other peers (Antle, 2003).

Academic prototypes and projects tend to afford more creativity (Cassell, 2008; Gottel, 2011). Many approaches make use of physical elements to trigger the process of story-making. For example, MIT’s StoryMat records and recalls children’s voices as they play with stuffed animals on a colorful, story-evoking, quilt (Cassell, Ryokai, 2001). Other approaches, like SAGE (Bers, Cassell, 1998) and PET (Druin et al., 1999), integrate tangible elements (like stuffed animals) into the technology-enhanced storytelling process. StoryRoom also adopts a physical approach by providing kids with room-sized interactive storytelling spaces where they share a theatrical experience (Alborzi et al., 2000).

More recent approaches exploit the affordances of tablet PCs technology. ShadowStory, for example, is a digital storytelling system inspired by traditional Chinese shadow puppetry. Using a Tablet PC children at primary school level create digital characters and perform live stories together on a projection screen (Lu et al., 2011). TinkRBook by MIT also makes use of tablets; aimed at very young children (pre-schoolers), it supports “storytelling behavior” since kids (by touching the interface) can modify the unfolding of the story (Chang, Breazeal, 2011).

Other approaches provide children with online tools for supporting the story creation process. A recent development is G-Flash, an authoring tool for primary school children that supports story creation using illustrated flashcards, with characters and scenarios (Jumail Rambli, Sulaiman, 2011). Wayang is another online authoring platform, developed by the dimeb Research Group of Bremen University, meant to allow students to express their cultural diversity. Children create either individual or collaborative stories by using digital puppets (Widjajanto et al., 2008).
Virtual environments are also used as “places” in which the stories can unfold. For example PUPPET is an autonomous agents-populated virtual environment where children play multiple roles in creating narratives (Marshall, Rogers, Scaife, 2002).

Collaborative storytelling has also been explored, but mostly at experimental level (Gottel, 2011). MOOSE crossing, for example, allows kids to cooperatively design and build objects and characters in a virtual space (Bruckman, 1997). The FaTe project allows very young kids (aged 5 to 8) to develop stories together in a shared 3D environment (Garzotto, Forfori, 2006). ToonTastic is a tool, still in its beta phase, meant to enable children to collaboratively create a story using an interactive, multiple-pen display (Russell, 2010). Digital drawing, especially if collaborative, has also been considered a form of storytelling, like in the KidPad project (Druin et al., 1997; Benford et al., 2000; Hourcade et al., 2002). Fails et alii (2010) present a system where children on a school trip collaboratively create a story using mobile applications.

Eventually, CBC (Canadian Broadcasting Corporation) 4Kids’s StoryBuilder is one of the rare examples of large-scale exploitation of a digital storytelling system. Children can create multimedia comix-style stories, based on the typical mechanism of “add-a-sentence-to-a-story”. They can then save their stories in an online personal space and also publish them and share them with friends, via email (Antle, 2003).

With respect to the above tools and programs, we can say that the uniqueness of PoliCultura (and its authoring environment) lies in the blend of the following aspects: (1) the class as a whole (not individual students) collaboratively creates the story; (2) the tool is not aimed at very young users exclusively but encompasses all age-ranges, from pre-schoolers to adults; (3) the creation process mingles formal and informal education: the whole work is coordinated by the teacher but many of the activities are performed out of the school following the students’ own initiatives; (4) there is always an educational goal, i.e. stories are never told/created for the sheer pleasure of telling (though they do give pleasure and satisfaction to their authors); (5) the technical tool is (very) easy and still the result is “technologically” surprising for the students, their teachers, their families etc., and provides a strong sense of accomplishment; (6) the final result undergoes a competition and it is then made public, providing additional motivations for the authors.

The PoliCultura initiative

In 2006, HOC-LAB, a laboratory at the DEI (Department of Electronics and Information) of Politecnico di Milano, developed the first version of a tool and an overall approach to multimedia production which was baptized “Instant Multimedia” (Di Blas et al., 2007). The rallying call was simplicity: non-tech savvy authors had to be enabled to develop, quickly and at low cost, multimedia presentations (with images, videos, audios, texts etc.). HOC-LAB had already gathered at that time a many-years’ experience in the field of education, with schools of all grades. The new tool was used for launching a new initiative: a competition for multimedia productions. In the first year, the competition was open to high schools only (grades 8-12 roughly). Participation was open to classes as a whole, not to individual students, to foster collaboration (Di Blas et al., 2010). During the first year a pilot project was run in a primary school in Milan (Italy). It was successful, so the following year, 2007, participation was open to primary schools and junior high schools too. In 2007 a preschool spontaneously took part in the program (Di Blas, Boretti, 2009); again, the result was positive, so from 2008 PoliCultura came to include all school grades.

If we consider that the authors of this paper use the authoring tool in higher education too (at University of Italian Switzerland and at Politecnico di Milano), we can say that PoliCultura is an example of “one size fits all” initiative.

Organization

Every year, the call for participation is disseminated through a number of channels.
- The local branches of the Italian Ministry for Education, at regional and at district’s levels
- More than 40 websites and blogs discussing education-related issues
- HOC-LAB’s mailing lists of teachers
- Direct mailing to all public and private Italian schools
- Public regional institutes for research on education
The call is issued in October and it is reiterated once a month, from November to January. The work must be delivered in March. A jury selects the best works according to the main competition’s categories, which correspond to the Italian school levels:

- PoliCultura First. For preschool, with students aged between 3 and 5
- PoliCultura Kids. For primary schools, with students aged between 6 and 10
- PoliCultura Junior. For Junior High School, with students aged between 11 to 13
- PoliCultura Senior. For High School, with students aged between 14 and 18

Participants can decide the subject of the story quite freely, provided it is somehow related to school’s curricula. In addition, some special tracks are proposed, also thanks to the sponsors that have shown interest in the initiative over the years. Table 1 shows the list of tracks for year 2011-12.

<table>
<thead>
<tr>
<th>Cultural heritage</th>
<th>Describe a monument, place, church of your territory, or describe your visit to a cultural institution (e.g., a museum). What are your impressions?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your roots</td>
<td>Describe the history, folklore, traditions of your territory. Make people curious about it!</td>
</tr>
<tr>
<td>School trip</td>
<td>Describe a particularly interesting school trip.</td>
</tr>
<tr>
<td>School ‘experience’</td>
<td>Describe an interesting educational experience: a lab, a research, a project…</td>
</tr>
<tr>
<td>Your Resurgence (the Italian unification movement)</td>
<td>Describe notable historical figures, places, events from your territory related to the Resurgence. What evidences are there of this important historical period? How much do people, and youngsters in particular, know about it today?</td>
</tr>
</tbody>
</table>
| Food and Nutrition | What is your relationship with food? What do you eat (at home, at school)? How do you choose your food? What do you know about the way vegetables are grown or animals are bred?  
**Sponsorship: Anagrhythos, National Association for Organic Agriculture** |
| Science and technology | Tell the story of your “discovery” of science and of your relation with technology. How has technology changed the way you live?  
**Sponsorship: Science and Technology Museum of Florence.** |
| Top 5 things to see and do in your town | Entice potential visitors to come and see the best of the place where you live!  
**Sponsorship: Venere, the online booking system of Expedia.** |
| Free choice!      | Tell us about whatever you feel interesting to tell!                                                                                                                                          |

Participants have three months circa to complete the work. In order to create the multimedia story, they are given free access to an authoring tool developed by HOC-LAB, 1001stories, a free web service that does not require any software installation. In April-May a jury composed of HOC-LAB communication experts, representatives of the sponsors and school teachers decides on the short list of finalists and the winners. The awards ceremony takes place in June at Politecnico’s premises: finalists from all over Italy are welcomed by authorities both from Politecnico and from the Italian ministry of Education, which officially acknowledges the initiative.
A story with 1001stories

A story done with 1001stories is composed by an audio commentary, complemented by a slideshow of images and videos plus text (the audio transcript), visible on demand (figure 3).

![Figure 3. The web version of a multimedia story](image)

A story is made up of a set of content elements, composed of audio, slideshow and text, each lasting one, two minutes on average. These elements are what Alexander and Levine define as “micro-content”, i.e., small chunks of content that contain an idea or concept, one of the characterizing aspects of the web 2.0 storytelling (Alexander,
Levine, 2008). These elements can be organized according to two, pre-defined, information architectures: (1) a two-levels hierarchical schema, with a set of topics (usually between 5 to 8) each with a set of sub-topics; (2) a linear schema, with a simple sequence of topics. Figure 4 shows an example of hierarchical schema: the set of topics and subtopics of “Padua Pocket”, a story done by a high school class from Padua, the beautiful Italian city chosen by Shakespeare as setting for “The taming of the shrew”. A class of foreign students is visiting the Padua’s school: to help them go around and enjoy the town’s highlights, the students have figured out a mobile “pocket” guide, complemented by interesting features like a handbook of Padua’s idioms and an interview with the town’s mayor.

Figure 4 shows an example of linear schema: the set of topics and subtopics of “Padua Pocket”, a story done by a high school class from Padua, the beautiful Italian city chosen by Shakespeare as setting for “The taming of the shrew”. A class of foreign students is visiting the Padua’s school: to help them go around and enjoy the town’s highlights, the students have figured out a mobile “pocket” guide, complemented by interesting features like a handbook of Padua’s idioms and an interview with the town’s mayor.

Figure 5, instead, shows an example of linear story (by a Swiss junior-high school): students go on a trip in a wood, looking for “erratic boulders” left over by glaciers and carved by primitive folks.

Figure 5. Example of linear information architecture (“Erratic boulders of Gandria”).

Obviously, the first schema requires more effort in terms of content organization and production, while the second is less demanding. Participants can freely choose between the two, according to how much time they mean to devote to the story-making activity.

Interaction possibilities vary according to the device. For example, in the web version, the user can choose between an automatic fruition of the story (either the sequence of macro topic or the sequence of all the topics and sub-topics) or free choice of topics and sub-topics. On iPhone or iPod instead, playlists are created, grouping for example all the main topics or a single topic and its sub-topics.

How to create a story

Participants work by themselves, without any form of tutoring. They are provided with a short manual (12 pages) where our own experience with creating stories using 1001stories is distilled. The manual suggests the steps for creating a multimedia story:

1. **Topic selection and gathering of the material**
   Generally speaking, the teacher puts forth a proposal for a topic; according to the students’ age, the topic is accepted or put under discussion. Sometimes the topic is something new; other times, the topic is related to some form of activity the class has already performed or is involved in (like for example a research, a project, a school trip…). In any case, the raw material for the narrative has to be gathered. If it is about something already known, it is a matter of selection, typically, from a huge heap of documentations. If it is about something new, then sources must be found. External actors get often involved, like for example the families, the local museum, experts, authorities, etc.

2. **Content organization**
   The information architecture is sketched, according to one of the proposed schemas: either the hierarchical or the linear one (figures 4 and 5). The titles of topics and sub-topics are defined. Students are usually organized into groups.

3. **Content preparation**
   The texts are written, synthesizing the various sources. The suggested length for each piece of content is between 120 and 200 words, corresponding to one minute of audio. Once the texts are finished, the audios are recorded. The visual communication is prepared: pictures, either taken by the students themselves, or found on the internet, scanned drawings, power point slides, etc. audios are recorded (sometimes with background music).

4. **First version**
   Audios, images and texts are uploaded in the authoring system and the first version is generated. It must be noted that nothing prevents uploading the single pieces of content (e.g., some images, some texts…) even before they are all ready.

5. **Evaluation**
   The first version is analyzed thoroughly. The most important check is whether it works from a communication point of view. The ‘evaluators’ (the students themselves) have to try all possible forms of interaction. For example, she they have to try whether the sequence of main topics makes sense as playlist.
6. **Final version**

Once all the flaws are found and amended, the final version is ready. 1001stories then generates automatically the various versions for the different channels.

The above steps are quite similar to the four stages that the Instructional Technology Department of the University of Houston (University of Houston, 2009) suggests to create a digital story. At stage one, the storytellers select the topic and gather the resources; at stage two, they select and organize the material; at stage three, they finalize the story in digital form and at stage four they present it to a reviewer to get feedback.

**The authoring tool**

Multimedia stories are created using the 1001stories environment. The current version is a highly sophisticated system that allows creating applications for a number of “channels”: web for PC, web for mobile, podcasts, iPad (and tablets in general), multi-touch tables, YouTube, standard phone (with audio-only content), etc. 1001stories has 3 main components: (1) an authoring environment where the various pieces of content are authored; (2) a number of generation engines that generate the proper information architecture, organizing the content items in a structure suitable for interactive usage; the information architecture is described via XML, according to a set of XML-schemas; (3) a number of delivery engines, implementing the various interactive formats over various platforms. The technical environment’s complexity is all internal: the final user finds it extremely easy to use. The average time for managing its basic commands is 20 minutes, for a staff of non-technical people.

**A mosaic of stories: the PoliCultura portal**

All the stories are gathered in a brand new, state of the art, portal (www.policultura.it/portal/). An advanced exploratory application allows users to access the stories using different canvases, highlighting different features, selecting different facets, etc. (fig. 6). Continuous feedback provides the user with information about how the different facets are correlated (e.g., the discipline and the type of learning experience, the specific subject and the level of schooling). Facets can also be correlated with words’ clouds derived from the abstracts (in English and Italian) of the stories. Other facets concern more pedagogical aspects. Users can drop their comments, though they cannot rate the stories, for obvious pedagogical reasons. PoliCultura-portal is a very large collection of students-generated multimedia content.

![Figure 6. The PoliCultura portal, two views: thumbnails (left) and mosaic (right).](image)

**Beyond the school’s boundaries**

The stories published on the portal, at September 2012, are 697; in June 2012 more than 150 additional narratives were collected, and they will be published by October 2012. Narratives are related to different school activities:

- 35% are about a visit to a cultural institution;
- 41% benefit from an expert’s contribution (e.g. a museum’s curator);
- 23% explicitly involve the contribution of students’ relatives (e.g. a grandmother);
- 10% involve in some way or another the local community (e.g. the city major, a local institution);
• 14% are about an activity outside of the school premises (e.g. a school trip);
• 29% imply (collaborative) working of the students and the teachers from home, beyond school hours.

So, in a large number of cases, storytelling implies activities outside the school’s premises and beyond the school hours. This latter aspect is facilitated by the fact that digital content (texts, images, audios…) can be prepared anywhere and at any time and that the tool (1001stories) is a free web service that students can access from home (if the teacher shares with them the account, which is often the case, especially at high-school level).

Table 2 shows some stories which exemplify the above aspects.

<table>
<thead>
<tr>
<th>Table 2. Examples of stories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Roman Milan</strong> – primary school, 2006</td>
</tr>
<tr>
<td>A primary school goes to visit the local archeological museum, than roams the town looking for evidences of its Roman glorious past. Children are divided into groups, each in charge of a specific topic (the baths, the theater…). They interview experts, take pictures, make and scan drawings to illustrate their journey into the past.</td>
</tr>
<tr>
<td><strong>A guide for students who are tired of sitting at their desks</strong> – junior high school, 2011</td>
</tr>
<tr>
<td>This story is about a cultural visit to the Romanesque cathedral and the museum of Assisi (a small town in central Italy). The pedagogical implementation of the experience was quite sophisticated: the students are divided into groups before the visit, they are in charge of taking interviews with the museum’s curator and an art-historian during the visit, they get back to school to elaborate the materials and then spontaneously go back again to the cathedral and the museum to gather additional materials. The result is a rich (in terms of content) and engaging guide to the famous cultural venue.</td>
</tr>
<tr>
<td><strong>Let’s go to school with grandma Daisy</strong> – primary school, 2011</td>
</tr>
<tr>
<td>How was school at our grandparents’ time? Grandma Daisy plays the expert and takes the kids back in time, showing how different things were when she was the kids’age.</td>
</tr>
<tr>
<td><strong>Padua Pocket</strong> – High School, 2010</td>
</tr>
<tr>
<td>This story is by definition “beyond the school’s boundaries”: it is a mobile guide for the city of Padua, done for a visiting class of foreign peers. “This multimedia guide is called ‘Padua pocket’. You can download it on your MP3 player. We will tell you about our town’s highlights, Hope you enjoy it!” Moreover, the final part of the story is an interview to the city major.</td>
</tr>
<tr>
<td><strong>The Erratic Boulders of Gandria</strong> (Switzerland) – junior high school, 2011</td>
</tr>
<tr>
<td>A Swiss class undergoes an unexpected adventure during a school trip: some of the kids get “lost”, maybe kidnapped by spirits of the woods. The story develops according to a script where each kid plays a role and it is complemented by pictures taken by the students themselves during the trip.</td>
</tr>
<tr>
<td><strong>From the “Nabucco” to the Mameli Anthem</strong> – 3 primary school classes, 2011</td>
</tr>
<tr>
<td>Three classes remotely located cooperate to tell the story of the two (eventually one) Italian anthems. They state in the introduction that “new technologies offer schools the opportunity to innovate and renew their teaching/learning environment, exploiting new ways of communication, like social network and skype, in a web 2.0 approach.”</td>
</tr>
</tbody>
</table>
Evaluation

PoliCultura is an ICT-based program that not only engages students (as it may be expected) but also provides them with substantial educational benefits. It can therefore be embedded into school’s curricula.

Every year, the educational impact of PoliCultura is evaluated through a number of means: online surveys, focus groups (during the celebration day of the competition), Skype interviews to the teachers as well as the analysis of the students’ stories. We do not collect data from students directly (which would be easily doable) for several reasons: first of all, it is normal for teachers to assess their students’ learning rather than for students to perform self-assessments; therefore we think that teachers can be quite reliable, and more reliable than students themselves. In addition, we want teachers to be co-designers of the experience and therefore we think that it is important that they pay attention to the learning process of their students. Finally, we think that it is quite likely that, with good teachers, the two evaluations are not so different. Our previous experience, in fact, with another large-scale set of programs (based on MUVEs), with more than 9,000 students involved, indicated that students’ self-assessment about benefits follows quite precisely the same curve of their teachers’ assessment, with only 10% circa less positive (Di Blas, Poggi, 2007). The small difference is likely due to the fact that when teachers evaluate the performance of the class, they tend to evaluate the “average”, implicitly giving less relevance to small groups of disaffected students. These less performing students provide, in general, negative evaluations that account for the small delta.

Almost all the teachers who take part in PoliCultura fill in the online survey. There are two reasons for the high response-rate: first of all, our staff is very active in pursuing teachers and “persuading them”. Secondly, filling in the survey is a pre-condition for actually taking part in the competition (some teachers complete the work but do not take part in the competition).

Data show that a number of benefits are achieved, ranging from increased knowledge of the subject at stake to improved social skills (group work) and attitudes. These benefits are widely discussed elsewhere (Di Blas, Paolini, 2012); for the sake of this paper’s argument, we will analyze only the data which can be put in relation to the “going beyond the school’s boundaries” aspect.

First of all, interacting with external institutions and experts is one of the elements that foster achievements in terms of increased knowledge (fig. 7). Students get a first-hand experience of works of art, archeological findings, exhibits introduced to them by an expert on the subject. Since they know they will have to build a “story” on top of the experience, they do not act as passive listeners but rather take on a quite active role, as interviewers. They get set before the visit, oftentimes preparing their interview’s schema and going through preparatory materials. During the visit (or when meeting the expert) they manage to get the best out of the experience. Of course this interaction is not the only reason behind increased knowledge: transcribing the interviews, synthesizing the materials, making additional researches through the internet and libraries do add to this benefit. But the real-world experience is surely the main trigger of this positive process.

![Cognitive benefits](image_url)

*Figure 7. Teachers’ evaluation of the students’ achievements of cognitive benefits (127 teachers, year 2011)*
Many of the factors that determine the high level of motivation participants to PoliCultura experience are related to the informal education aspects of the program. The competition at national level is for sure one of the strongest. Other factors are the involvement of out-of-the-school actors (like the families and the local community), with whom the final story will be shared, and the public visibility of the result (all the stories, as we said, are published in the online portal). Data show that motivation is quite high even for the normally disaffected and underperforming students (fig. 8). Again, the external context is a key factor.

A teacher (from pre-school) reports: “to motivate the kids, we told them they were about to create a multimedia book that their parents and many other people would read […] and actually, when the work was finished, it was shown in the local theater and even the town’s major came to see it” (Di Blas, Boretti, 2009).

![Figure 8. Teachers’ evaluation of the students’ motivation (127 teachers, year 2011)](image)

Eventually students’ improvement in communication skills and also in “media literacy” (i.e., the ability to decode, understand, and write through, and with, all forms of media, in specific with text, audio and images in multimedia form), is quite substantial (fig. 9). Similar results are reported by Xu et al. (2011) on the efficacy of digital storytelling to teach writing skills (see also Gakhar & Thompson, 2007 and Robin, 2008).

A teacher reports: “since texts had to be short and ‘to the point’, students have developed good analysis and synthesis abilities, which in normal school activities are not triggered”.

![Figure 9. Teachers’ evaluation of the communication benefits (127 teachers, year 2011)](image)
Conclusions

PoliCultura shows that it is possible to bring innovation in a conservative environment like the school, blending formal and informal education while promoting out-of-the-classroom activities. Data show that in doing so students achieve substantial educational benefits, many of which specifically related to this approach: the vast majority of participants involved other subjects like families, cultural institutions, museums, local associations of various kinds, local tourist bureaus, politicians, experts of various subjects, book writers, and even old immigrants. Students roamed the territory using various kinds of devices, like their cell phones or digital cameras, to gather content that was then refined in the classroom.

There are many reasons why educational activities should not be confined to the standard space-time constraints of classroom and school hours. The most important reason is probably social pressure: the notion of being “always connected” is pervasive in the society, and it is taking ground at school too. The danger is that the very foundations of the school systems can be shaken by this evolution. PoliCultura shows a good and working compromise: schools can get out of their boundaries (in several ways) still retaining control over the learning process and ensuring that sound educational benefits are gained. As data show, is not just matter of “fun and engagement”; it is also matter of cognitive and behavioral improvements.

Our plans for the future include widening the range of narrative formats, and also suggesting (and supporting) different workflows for narrative creation. From a research point of view, we intend to better investigate the correlation between the “implementation strategy” (class organization, activities scheduling, groups’ organization, interaction with external actors…) and the educational benefits, as they are measured by the teachers. From an organization point of view, we plan to scale up PoliCultura to international level, European first and then worldwide.

Acknowledgments

We thank the people of the HOC-LAB staff who passionately work for making PoliCultura a success every year, especially Elena Maccari, PoliCultura’s project manager. This work is partially supported by National Project L4A (“Learning for All”) Grant Num RBNE07CPX 001.
References


Investigating the Activities of Children toward a Smart Storytelling Toy

Nuri Kara1*, Cansu Cigdem Aydin2 and Kursat Cagiltay1

1Department of Computer Education and Instructional Technology, Middle East Technical University, Turkey // 2Department of Computer Engineering, Atilim University, Turkey // nkara@metu.edu.tr // cansu@atilim.edu.tr // kursat@metu.edu.tr

*Corresponding author

ABSTRACT
This paper introduces StoryTech, a smart storytelling toy that features a virtual space, which includes computer-based graphics and characters, and a real space, which includes plush toys, background cards, and a communication interface. When children put real objects on the receiver panel, the computer program shows related backgrounds and characters on the screen. StoryTech encourages children to produce their own stories. Empirical research with a total of 90 children indicated that StoryTech contributed to narrative activities of children and made a positive impact on their creativity. The paper also focuses on the attributes, design, and development process of StoryTech.

Keywords
Storytelling, smart toy, mixed reality, children

Introduction
Smart technologies have been adapted to several systems for the purpose of integrating technological innovations. These technologies have more capabilities to meet humans’ needs than previous traditional ones. For instance, presenting multimedia content and providing instant feedback can be effectively achieved with these smart technologies. Toys have also gained digital characteristic with these smart technologies, such as computers, software, and digital materials. These newly developed toys can be called smart toys. The advantage offered by these smart toys is that they provide a rich play environment in terms of multimedia use and interactive usability by integrating the digital medium with traditional toys. As Lampe and Hinske (2007) stated, enriching traditional toys by adding multimedia content can increase the quality of fantasy play and provide a more creative environment for children. Smart toys can provide several alternatives for children, such as supporting collaborative play, active reaction, and imaginative play.

Fantasy play is of great importance in the lives of children because it can reflect their emotions and imagination freely. As Trionfi and Reese (2009) reported, there is a positive relationship between fantasy play and the storytelling abilities of children. Similar to this view, Cassell and Ryokai (2001) stated that fantasy play enables children to reveal their inner world possibilities without facing difficulties. Fantasy play can be viewed as combining physical reality with children’s imaginative world. From this perspective, smart toys that combine physical reality with virtual reality can be a good option for children to engage in fantasy play, including storytelling, role-playing, etc. These smart toys can also generally be called computationally augmented toys (Bers & Cassell, 1998; Glos & Cassell, 1997), digitally augmented physical spaces, digitally enhanced physical spaces (Price & Rogers, 2004), and other similar terms. The advantage offered by these smart toys is that they provide mixed reality by combining physical and virtual realities (Stapleton, Hughes, & Moshell, 2002; Lampe & Hinske, 2007). This paper introduces a smart storytelling toy called StoryTech. It is used to improve children’s creativity by helping them to create rich stories. StoryTech can be used as a part of formal learning activities under the supervision of teachers. It can also be used as an informal learning tool. Children can also create and record stories by themselves. The researchers investigated the impact of StoryTech on children’s narrative activities and creativity.

Related Work
There have been several smart storytelling toy projects in the field: Rosebud is an interactive system that links children’s stories to their toys (Glos & Cassell, 1997). Rosebud is a keyboard- and desktop-dependent system that requires technical collaboration with a computer. StoryMat is a computationally augmented space that encourages children’s storytelling and story-listening activities (Cassell & Ryokai, 2001). StoryMat provides a play space for children to tell their own stories by recording and recalling their stories. Compared with Rosebud, stuffed animals are
the main characters of the system rather than a desktop computer. Dolltoy is a toy that allows children to record their stories and hear them back by capturing the gestures and speech of a child (Vaucelle & Jehan, 2002). StoryToy is an augmented toy that tells different stories based on the play actions of children with physical toy objects related to a farm (Fontijn & Mendels, 2005). Different from StoryMat or Rosebud, the aim of StoryToy is not only to encourage the children to tell their own stories but also to create a traditional play environment including a predefined story content (Fontijn & Mendels, 2005). One of the common points of most of these smart toys is to use simple low-tech stuffed animals to provide a natural feel while playing. Another common point is to lead children to a storytelling activity by combining stuffed characters with technological devices. The same characteristics are also valid for StoryTech. Compared with previous smart toys, the advantage of StoryTech, on the one hand, is that it does not only provide a free play space for children but also encourages their storytelling activities. On the other hand, this study also provides a detailed and large-scale user study with a total of 90 children compared with related previous studies.

**Storytelling**

Storytelling can be viewed as an activity that allows people to share their experiences, thoughts, or imagination with verbal statements. As Broström (2002) stated, “Storytelling is a genuine human way to express experiences, thoughts and feelings” (p. 86). Especially in children’s fantasy world, storytelling has a great role since it provides a tool to reflect the children’s inner feelings. Storytelling that includes visual and oral expressions of stories in a physical space can deeply enhance creativity and imagination (Sugimoto, 2011). This activity also enables children to use language without feeling any restriction or difficulty. Telling stories contribute to children’s language learning by helping them choose the appropriate vocabulary and structure the stories in a coherent way (Loizou, Kyriakides, & Hadjicharalambous, 2011). According to Cassell and Ryokai (2001), children have control in terms of using language in storytelling. Since storytelling does not require any systematic work for children, it provides them an enjoyable environment. Similar to this view, Ryokai, Vaucelle, and Cassell (2003) pointed out that storytelling is an ideal activity for children since writing tasks are not necessary.

Storytelling can also be considered an important factor in terms of cognitive development. According to Loizou, Kyriakides, and Hadjicharalambous (2011), stories contribute to the cognitive development of children. Since children take different roles (such as narrator, character, or themselves while telling their stories), they need to be cognitively active in this process. As Cassell and Ryokai (2001) emphasized, the ability to shift between roles provides a cognitive activity for children. Additionally, linking experiences or imagination with the toys or objects in an existing environment also stimulates the cognitive abilities of children. For instance, telling a story by adapting a lived experience of an animal character is the result of using cognitive processes.

**Peer Collaboration**

Peer collaboration enables children to become socialized by sharing their activities with other children. The important advantages of collaborative work include not only socializing but also developing several skills. As Ryokai and Cassell (1999) emphasized, collaboration allows children to improve their language and narrative skills. During the collaboration process, children also have a greater role in activities. More collaboration provides more opportunities to share in terms of ideas, suggestions, and experiences (Price & Rogers, 2004). Storytelling can be considered a suitable activity for children to create a collaborative environment. Children can share their imagination or experiences with their peers by using verbal statements. Collaborative storytelling also provides a way to receive immediate feedback and interaction in a physical space (Liu, Liu, Chen, Lin, & Chen, 2011). According to Ryokai, Vaucelle, and Cassell (2003), storytelling is an important factor in peer collaboration since it increases the awareness of children toward the language skills.

**StoryTech**

The aim of starting the StoryTech project was to create a dynamic storytelling toy for children by providing a mixed-reality environment. This toy can be used as a part of formal (in-class) learning setting or it can also be used individually as an informal learning tool. In this study, children used it individually and created their own stories without any supervision. The design and development period of the project lasted approximately 10 months.
analysis process, interviews with academicians from both early childhood education and instructional technology fields were carried out for several purposes, such as to determine the needs and preferences of children, to determine the necessity of the toy as an instructional technology, and to better understand expectations from both fields. Interviews with toy industry experts were also conducted to determine the types of preferred toys in the industry and to understand their approaches to the smart toy concept. Previous smart toy projects were analyzed to determine a road map for the development of a new smart toy. Literature was also reviewed to strengthen the pedagogical side of the project. In the design process, four different smart toy ideas were offered. Additionally, paper-based templates and storyboards were composed. The last step of the design process was to decide on one of the four project ideas. During the decision-making process, opinions of academicians and experts were also solicited. In the development process, technical devices like the RFID and RFID reader required to develop the decided project were acquired. Afterward, the scenes were created, and action script codes were written in Adobe Flash program in line with the storyboard developed in the design process. The RFID system was also combined with the Adobe Flash environment in this process. In the testing process, a user study was carried out with a total of 90 children in different kindergartens after securing the necessary permits from authorities. Figure 1 below also summarizes the whole process.

### Figure 1. Timeline of the StoryTech project development process

<table>
<thead>
<tr>
<th>Analysis (3 months)</th>
<th>Design (2 months)</th>
<th>Development (2 months)</th>
<th>Testing (3 months)</th>
</tr>
</thead>
</table>
| • Interview with academicians from early childhood education and instructional technology<br>• Interview with toy industry experts<br>• Review previous smart toy projects<br>• Literature review<br>• Offer four different smart toy ideas<br>• Compose templates and storyboards of each idea<br>• Decide on one of the four ideas to develop<br>• Provide technical devices (e.g., RFID and RFID reader) and stuffed animal characters<br>• Draw the environment and characters in Adobe Flash program<br>• Write action script codes<br>• Combine RFID technology with the Adobe Flash program<br>• Select kindergartens to conduct the user study<br>• Secure the necessary permits from authorities<br>• Conduct the study with the target group

Structure of the StoryTech

This child-driven smart toy consists of three main parts (Figure 2). The first part includes story objects that contain four plush toys (polar bear, rabbit, turtle, and sheep) and four background cards (night, farm, polar, and cottage backgrounds). The background cards consist of pictures of an environment where a story takes place. The second part uses a communication interface that includes an RFID (radio-frequency identification) system to transfer the data to the computer. The story object will then be shown in Flash animation. The last part consists of a computer that includes a Flash application developed with Adobe Flash.

The plush toys and background cards include RFID tags. While these tags are placed on the bottom of the plush toys, these tags are combined with background pictures as cards. These tags give a unique code to the related story object, and this code is transferred to the Adobe Flash software via the RFID reader. That is, when the related story object is
put on the RFID reader, the unique code of the object, given by the RFID tag, is transferred to the computer via the RFID reader, and this code is used in Adobe Flash to create a virtual character of the related object.

Two different Flash modules are in the smart toy. The aim of the first module is to make the children ready for the second (storytelling) module. In first module, only the prerecorded story of the rabbit and turtle is narrated, and children are expected to put only the appropriate character (rabbit or turtle) on the interface in line with the narration. When a child puts the right one on the interface (RFID reader), the related story object appears on the screen. However, when the child puts the wrong one on the interface, the story object does not appear on the screen until the child puts the right one. During the interaction of story objects with Flash animation, the sound and visual effects are used to attract the attention of children. This module is called scaffolding module because it helps children learn the function of StoryTech. After finishing the scaffolding module of StoryTech, the children are expected to start with the second module of StoryTech. The aim of the second (storytelling) module is to encourage children to tell their original stories. In this module, children can put any story objects (plush toys and background cards) on the interface, and they can create their own stories. When they put any of the plush toys on the interface, a related animal character appears on the screen as an animation; when they put it again, the character disappears from the screen. Similarly,
they can change the environment on the screen by putting different background cards on the interface. Children are also led to tell their stories with the narrator present in the module. This narrator describes the environment based on the background card placed on the communication interface and reminds children to continue their story. The Adobe Flash environment presenting the virtual characters of the story objects placed on the communication interface with multimedia features and narrations supports the story creation of children. The child should handle and place the plush toy or background card properly with his/her story on the communication interface and continue his/her story by changing the plush toys and background cards. There is also no difference in the operation of the system when two children, instead of just one, play with StoryTech. In a dyads play, the child should follow his/her peer’s story and continue to tell the story by changing story objects on the communication interface. Figure 3 shows photos of children playing with StoryTech.

Methodology

Participants and Sampling

Ninety preschool children selected from five different local kindergartens in Ankara were included in the study. The two-stage sampling was applied as the sampling method. Five different local kindergartens in Ankara were purposefully selected by the researchers. Then 18 children including six children per age-group (4, 5, and 6 years old) were randomly included in the study. In Table 1, the demographics of the children are presented.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>47</td>
<td>52.22</td>
</tr>
<tr>
<td>Male</td>
<td>43</td>
<td>47.78</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 years old</td>
<td>30</td>
<td>33.33</td>
</tr>
<tr>
<td>5 years old</td>
<td>30</td>
<td>33.33</td>
</tr>
<tr>
<td>6 years old</td>
<td>30</td>
<td>33.33</td>
</tr>
</tbody>
</table>

Design

The design of the study was experimental design. Children were randomly assigned into one of four groups according to their ages. The name of the groups were the control group for single children, control group for dyads, StoryTech group for single children, and StoryTech group for dyads. Each group was also divided into three subgroups based on the age of the children. That is, each group had three subgroups: a control group including single children in the age 4 group, a control group including single children in the age 5 group, and a control group including single children in the age 6 group. There were five children for each single group and five dyads for each dyads group. In the single group, the children played alone; in the dyads group, the children played with another child.

Procedure

Permissions to study the children and record on video their play were secured before conducting the research in each kindergarten. A room suitable to hold the study was determined in each kindergarten, and the children were taken inside the room one by one. One of the teachers in each kindergarten accompanied the children to the room. A teacher was also in the room to help the children adapt to the study easily. There were two researchers, a teacher, and the participant in the room. One of two researchers used a video camera to record the children playing, and the other one explained the instructions to the child before he/she started playing. The “Once upon a time” starting sentence was used, and how to play with both the passive toy and StoryTech was explained in detail to each of the children. Since StoryTech had two parts (namely, the scaffolding part to adapt the children to the smart toy and the storytelling
part), additional explanations for the storytelling part were also given after finishing the first part. The children played with both the passive toy and StoryTech without time restriction.

Data Analysis Plan

The research questions of the study are as follows:
1. How does playing with StoryTech impact the children’s narrative activities?
2. How does playing with StoryTech impact the children’s creativity?

For this purpose, the data analyzed with this user study are as follows:
- The discourse patterns of the children in both StoryTech and control groups
- The number of imaginative objects produced during storytelling

To find the discourse patterns, talks included in the video recording were analyzed descriptively. Researchers first coded in MS Excel the speech of the representative child from each of the age-groups, and then the discourse of the representative child was prepared as a graphic with a timeline. To analyze the number of imaginative objects, the number of different real objects in the environment that children transformed into fantasy objects was counted (Sheldon & Rohleder, 1995; Cassell & Ryokai, 2001). The background cards were taken as real objects since the children mostly produced imaginative objects depending on the background cards rather than on the stuffed animals. Two-way ANOVA analyses were also applied to compare the groups. In the first analysis, the experiment group had two levels such as a control subgroup and StoryTech subgroup, and the group type had two levels such as the single subgroup and dyads subgroup, which were the independent variables. The goal of the first two-way analysis was to find whether there is a significant difference between these independent groups. In the second analysis, the experiment group had two levels such as the control subgroup and StoryTech subgroup, and the age-group factor had three levels such as those from age 4, age 5, and age 6, which were considered independent variables. The goal of this two-way ANOVA analyses was also to find whether there is a significant difference between these groups. The number of imaginative objects was considered the dependent variable for both analyses.

Limitations of the Study

Since age was also included in the analysis as an important factor, each group consisted of only five participants. Additionally, two children for the single age 4 group who played with the passive toy, two children for the single age 4 group who played with StoryTech, one dyad for the dyads age 4 group who played with the passive toy, and two dyads for the dyads age 4 group who played with StoryTech did not join the study because they felt uncomfortable. Additionally, the two-way ANOVA analyses were conducted rather than the three-way ANOVA. The first reason behind this choice is the limited number of children in each group. The second reason is to compare the groups and to decrease the margin of error. Although the events and actions would be good indicators of creativity in storytelling, only the imaginative objects were considered indices of creativity in this study.

Results

Children’s Story Patterns

In this part, the discourse patterns of children are presented as graphics. These discourse graphics are based on several activities such as the child playing, the child talking, the animation talking, and the support. The definitions of these activities are as follows:

Child playing: For the control group, this activity is related to moving around the plush toys on the background cards without telling stories. For the StoryTech group, this activity is related to only putting or lifting plush toys or background cards on the receiver panel without telling stories.
Child talking: This activity is related to talking that consists only in storytelling activity for both the control group and StoryTech group.

Animation talking: This activity is only valid for the StoryTech group. It is related to the narrations presented by the system in an Adobe Flash environment. In the first part of StoryTech, the narrator in the multimedia environment tells a story about a rabbit and turtle. In second part, the narrator describes environments that match with the background cards placed on the communication interface.

Support: This activity is related to interventions of the teacher or researcher. The support activity consists of both helping the child use the passive toy or StoryTech and helping the child construct a meaningful story. Although help was not solicited by the children most of the time, the teacher or researcher assisted the children when they have difficulty with the tool or telling a story.

**Discourse Patterns of Single Children**

Figure 4 shows a sample of a 6.58-minute session of children playing with passive toys from each of the three age-groups. Figure 4 indicates the frequency of children talking and the details of the talking procedures of the children during the storytelling activity. According to Figure 4, the duration of the storytelling activity of 4-year-old children was less than that of 5- and 6-year-old children. The graph also indicates, on the one hand, that children in the age 4 group received more support than children in the age 5 and age 6 groups in their limited storytelling activity, and the talking time of the 4-year-old children was at its lowest level. On the other hand, the 5-year-old children continued the storytelling activity more than the 4-year-old children. They mostly tended to tell their stories rather than only play with toys. The talking time of the 5-year-old children was also higher than that of the 4-year-old children. Based on the discourse patterns, the children who played with plush toys and background cards in the age 6 group told their stories throughout the 6.58-minute session. The graph also shows that 6-year-old children emphasized almost all components, such as playing, talking, and receiving support.

![Figure 4](image)

*Figure 4. Discourse patterns of children who play with the passive toy*

Figure 5 shows a sample 8.45-minute session of children who played with StoryTech from each of the three age-groups. Figure 5 indicates the frequency of children talking and the details of the talking procedures of the children during the storytelling activity. According to Figure 5, the duration of the storytelling activity of the 4-year-old children was less than that for the 5- and 6-year-old children. The figure also indicates that children in the age 4 group received more support than children in the age 5 and age 6 groups in their limited storytelling activity, and the
talking time of the 4-year-old children was at its lowest level. The figure shows that, on the one hand, the 4-year-old children mostly focused on playing rather than talking. On the other hand, the 5-year-old children continued the storytelling activity more than the 4-year-old children. The 5-year-old children mostly tended to tell their stories rather than play only. Based on the discourse patterns, children in the age 5 and age 6 groups showed similarity in terms of the talk time and play time with StoryTech. However, the 6-year-old children received less support than the 5-year-old children.

**Figure 5.** Discourse patterns of children who played with StoryTech

**Figure 6.** Discourse patterns of the dyads that played with the passive toy

*Comparison of Discourse Patterns of Single Children*

The findings of the discourse patterns indicate that single children from all age-groups who played with the smart toy spent more time in the storytelling activity than the single children from all age-groups who played with the passive toy. The 4-year-old children who played with the smart toy focused on play activity more than the 4-year-old
children who played with the passive toy. Findings also pointed out that the talk times of children who played with the smart toy in the age 5 and age 6 groups were higher than the children who played with the passive toy in the same age-groups. The children who played with the smart toy in the age 5 and age 6 groups spent a big amount of time telling a story rather than playing with the smart toy.

**Discourse Patterns of Dyads**

Figure 6 shows a sample 5.57-minute session of dyads that played with the passive toy from each of the three age-groups. Figure 6 indicates the frequency of dyads talking and the details of their talking procedures during the storytelling activity. According to Figure 6, the duration of the storytelling activity of the dyads in the age 4 and age 5 groups was less than that of dyads in the age 6 group. Additionally, both the 4- and 5-year-old dyads focused on playing more than the 6-year-old dyads. The 6-year-old dyads mostly emphasized on talking, and the second child was involved in storytelling activity more than the dyads in the age 4 and age 5 groups.

Figure 7 shows a sample 13.18-minute session of the dyads that played with the smart toy from each of the three age-groups. Figure 7 indicates the frequency of dyads talking and the details of the talking procedures of the children during the storytelling activity. According to Figure 7, the duration of the storytelling activity of the dyads in the age 4 group was less than that of the dyads in the age 5 and age 6 groups. The figure also indicates that the 4-year-old dyads received more support than the 5- and 6-year-old dyads in their limited storytelling activity. Additionally, the talk time of the 4-year-old dyads was at its lowest level. The figure shows that the 4-year-old dyads mostly focused on playing rather than talking. The 5-year-old dyads continued their storytelling activity more than the 4-year-old dyads. The 5-year-old dyads mostly tended to tell their stories rather than only play. Based on the discourse patterns, the 5-year-old and 6-year-old dyads showed similarity in terms of time of talking and playing with the smart toy. However, it can be stated that the second child was more active among the dyads in the age 6 group than the dyads in the age 5 group.

**Comparison of Discourse Patterns of Dyads**

The findings of the discourse patterns indicate that the dyads from all age-groups that played with the smart toy spent more time on the storytelling activity than the dyads from all the age-groups that played with the passive toys. The 4-year-old dyads who played with both the passive toy and smart toy did not give much emphasis on talking in a similar way. The 4-year-old dyads that played not only with the passive toy but also with smart toy had the least
storytelling duration. Findings also pointed out that the dyads who played with the smart toy in the age 5 and age 6 groups told their stories more than the counterparts who played with the passive toy. The talk times of the 5- and 6-year-old dyads that played with the smart toy were also higher than the same age-groups’ dyads that played with the passive toy.

**Number of Imaginative Objects**

Fantasy play leads children to enhance their creative, cognitive, and reflective skills (Schachtel, 1959; as cited in Cassell & Ryokai, 2001). Children can compose their own worlds by reflecting their inner thoughts in pretend activities. They might also be at a higher cognitive level in a pretend mode than in a non-pretend mode (Lillard, 1993). From this point of view, imaginative objects can provide an important clue to understand how deep the children use their cognitive functions during fantasy play. Cassell and Ryokai (2001) investigated the number of imaginative objects the children transformed from realistic objects in the play situation. For instance, when a child used a house as a police station, the police station was considered an imaginative object. Similarly, we investigated the variety of imaginative objects transformed from background cards as reality objects. Children produced imaginative objects in their talks during the storytelling activity. The reason behind the use of background cards rather than plush toys as reality objects is that these cards have more open-ended features to use in both the passive toy and StoryTech. Children can comment on the background scenes more than the plush toys.

**Imaginative Objects of Single Children Who Played with the Passive Toy**

In Table 2, a variety of imaginative objects transformed from the background cards during the storytelling activity is offered. According to Table 2, the 6-year-old children transformed more imaginative objects than the 4- and 5-year-olds. Children in the age 4 group transformed the least number of imaginative objects.

<table>
<thead>
<tr>
<th>Background cards</th>
<th>Age 4 group (3 children)</th>
<th>Age 5 group (5 children)</th>
<th>Age 6 group (5 children)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background 1</td>
<td>snowing</td>
<td>snowing</td>
<td>snowing</td>
</tr>
<tr>
<td>Background 2</td>
<td></td>
<td>farm</td>
<td>morning</td>
</tr>
<tr>
<td>Background 3</td>
<td></td>
<td>room</td>
<td>house</td>
</tr>
<tr>
<td>Background 4</td>
<td></td>
<td>night</td>
<td>evening</td>
</tr>
</tbody>
</table>

Children can comment on the background scenes more than the plush toys.
**Imaginative Objects of Single Children Who Played with StoryTech**

In Table 3, a variety of imaginative objects transformed from the background cards during the storytelling activity is offered. According to Table 3, the 6-year-old children transformed more imaginative objects than the 4- and 5-year-olds. Children in the age 4 group transformed the least number of imaginative objects. The findings in Table 3 also indicate that background 3 and background 4 provided the most imaginative objects for the 6-year-old children. For children in the age 5 group, the most imaginative object was produced using background 1.

<table>
<thead>
<tr>
<th>Background cards</th>
<th>Age 4 group (3 children)</th>
<th>Age 5 group (5 children)</th>
<th>Age 6 group (5 children)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background 1</td>
<td>snowing</td>
<td>a cold place</td>
<td>snowing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>snowing</td>
<td>house of the snowman</td>
</tr>
<tr>
<td></td>
<td></td>
<td>winter</td>
<td>house of the polar bear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>land of the snows</td>
<td>polar</td>
</tr>
<tr>
<td>Background 2</td>
<td>early morning</td>
<td>morning</td>
<td>morning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forest</td>
<td>outdoors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>summer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>garden</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>farm</td>
</tr>
<tr>
<td>Background 3</td>
<td>house</td>
<td>house of the animals</td>
<td>inside the cottage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inside the farm</td>
<td>cottage of the animals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>morning</td>
<td>morning</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>kitchen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>house</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a hot place</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mill</td>
</tr>
<tr>
<td>Background 4</td>
<td>night</td>
<td>evening</td>
<td>street</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night</td>
<td>midnight</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dark night</td>
<td>dark</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>the place for sleeping</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>evening</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>night</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>dark night</td>
</tr>
</tbody>
</table>

**Comparison of Imaginative Objects between Single Children Who Played with the Passive Toy and Smart Toy**

Findings clearly indicate that single children who played with the smart toy produced more imaginative objects than single children who played with the passive toy. While single children who played with the passive toy from all age-groups produced a total of 13 different imaginative objects, the number of imaginative objects was 40 for single children who played with StoryTech.

**Imaginative Objects of Dyads Who Played with the Passive Toy**

In Table 4, a variety of imaginative objects transformed from the background cards during the storytelling activity is offered. According to Table 4, the 6-year-old dyads transformed more imaginative objects than the 4- and 5-year-olds. Dyads in the age 5 group transformed the least number of imaginative objects.

**Imaginative Objects of Dyads Who Played with StoryTech**

In Table 5, a variety of imaginative objects transformed from the background cards during the storytelling activity is offered. According to Table 5, the 6-year-old dyads transformed more imaginative objects than the 4- and 5-year-olds. Dyads in the age 4 group transformed the least number of imaginative objects.
### Table 4. Imaginative Objects of Dyads Who Played with the Passive Toy

<table>
<thead>
<tr>
<th>Background cards</th>
<th>Age 4 group (4 dyads)</th>
<th>Age 5 group (5 dyads)</th>
<th>Age 6 group (5 dyads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background 1</td>
<td>snow</td>
<td>cold place</td>
<td>snow</td>
</tr>
<tr>
<td>Background 2</td>
<td>farm</td>
<td>garden</td>
<td></td>
</tr>
<tr>
<td>Background 3</td>
<td>box</td>
<td>inside of the house</td>
<td>house treasury room</td>
</tr>
<tr>
<td>Background 4</td>
<td>house</td>
<td>night</td>
<td>evening</td>
</tr>
</tbody>
</table>

### Table 5. Imaginative Objects of Dyads Who Played with StoryTech

<table>
<thead>
<tr>
<th>Background cards</th>
<th>Age 4 group (3 dyads)</th>
<th>Age 5 group (5 dyads)</th>
<th>Age 6 group (5 dyads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background 1</td>
<td>polar</td>
<td>the country of polar</td>
<td>polar</td>
</tr>
<tr>
<td></td>
<td>bears</td>
<td>bears</td>
<td>North Pole snowing</td>
</tr>
<tr>
<td>Background 2</td>
<td>morning</td>
<td>forest</td>
<td>morning</td>
</tr>
<tr>
<td></td>
<td>farm</td>
<td>summer</td>
<td>morning</td>
</tr>
<tr>
<td>Background 3</td>
<td>kitchen</td>
<td>inside the house</td>
<td>treasury home</td>
</tr>
<tr>
<td></td>
<td>house</td>
<td>room</td>
<td>treasury room</td>
</tr>
<tr>
<td></td>
<td>room</td>
<td>home</td>
<td>home</td>
</tr>
<tr>
<td>Background 4</td>
<td>dark</td>
<td>farm</td>
<td>night home</td>
</tr>
<tr>
<td></td>
<td>dark night</td>
<td>dark farm</td>
<td>dark farm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>evening</td>
</tr>
</tbody>
</table>
Findings clearly indicate that the dyads that played with the smart toy produced more imaginative objects than the dyads that played with the passive toy. While the dyads that played with the passive toy from all different age-groups produced a total of 12 different imaginative objects, the number of imaginative objects was 26 for the dyads who played with StoryTech.

**Two-Way ANOVA Results**

Before conducting the first two-way ANOVA analysis, the alpha level was selected, and the assumptions were checked. The selected alpha level was .01. Since the children in each group were independently observed, the first assumption related with independent observation was met. To check the normality assumption, skewness and kurtosis values and histograms were controlled. Since the skewness and kurtosis values of each group for all independent variables were between -2 and +2, this was the first reference point to meet the assumption. Histograms also showed that the normality assumption was met. For the homogeneity of variances assumption, Levene’s test value was controlled (Field, 2009). Since the test was significant, we concluded that the homogeneity of variances assumption was not met (p = .00). Hence, the alpha level was divided into two to meet the assumption. A new alpha level was determined as .005.

A 2x2 ANOVA was conducted to evaluate the effects of the experiment group (those who played with the smart toy or passive toy) and group type (single children or dyads) on producing the number of imaginative objects. The analysis indicated no significant interaction between the experiment and group type, $F(1,49) = 2.62, p = .11, \text{partial } \eta^2 = .05$, and no significant main effect for group type, $F(1,49) = 4.65, p = .04, \text{partial } \eta^2 = .09$. That is, there is no significant main difference between the single children and dyads in terms of the number of imaginative objects. However, there is a significant main effect for the experiment group, $F(1,49) = 38.68, p = .00, \text{partial } \eta^2 = .44$ (see Table 6). That is, there is a significant main difference between the experiment group who played with the smart toy ($M = 2.58$) and the control group who played with the passive toy ($M = 0.93$) in terms of the number of imaginative objects produced.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>35.964</td>
<td>1</td>
<td>35.964</td>
<td>38.679</td>
<td>.000</td>
<td>.441</td>
</tr>
<tr>
<td>Type</td>
<td>4.322</td>
<td>1</td>
<td>4.322</td>
<td>4.648</td>
<td>.036</td>
<td>.087</td>
</tr>
<tr>
<td>Experiment * type</td>
<td>2.431</td>
<td>1</td>
<td>2.431</td>
<td>2.615</td>
<td>.112</td>
<td>.051</td>
</tr>
<tr>
<td>Error</td>
<td>45.560</td>
<td>49</td>
<td>.930</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>88.302</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similar to the first two-way ANOVA analysis procedure, the alpha level was selected, and assumptions were checked before conducting the second two-way ANOVA analysis. The selected alpha level was .01. Since the children in each group were independently observed, the first assumption related with independent observation was met. To check the normality assumption, skewness and kurtosis values and histograms were controlled. Since the skewness and kurtosis values of each group for all the independent variables were between -2 and +2, this was the first reference point to meet the assumption. Histograms also showed that the normality assumption was met. For homogeneity of variances assumption, Levene’s test value was controlled (Field, 2009). Since the test was not significant, we concluded that the homogeneity of variances assumption was met (p = .011).

A 2x3 ANOVA was conducted to evaluate the effects of the experiment group (those who played with the smart toy or passive toy) and the age-group (age 4, age 5, and age 6) on the number of imaginative objects produced. The analysis indicated a significant interaction between the experiment and age-groups, $F(2,49) = 15.88, p = .00, \text{partial } \eta^2 = .40$; a significant main effect for experiment, $F(1,49) = 62.17, p = .00, \text{partial } \eta^2 = .57$; and a significant main effect for the age-group, $F(2,49) = 10.16, p = .00, \text{partial } \eta^2 = .51$ (see Table 7). Post hoc comparisons that use the Tukey HSD test indicated that the mean score for the age 6 group ($M = 2.50$) was significantly different from the age 4 group ($M = 1.00$) and the age 5 group ($M = 1.45$). However, the mean score for the age 5 group was not significantly different from the age 4 group.
Table 7. Two-way ANOVA Analysis for the Experiment and Age-group

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>26.06</td>
<td>1</td>
<td>26.06</td>
<td>62.17</td>
<td>.00</td>
<td>.57</td>
</tr>
<tr>
<td>Age-group</td>
<td>20.31</td>
<td>2</td>
<td>10.16</td>
<td>24.23</td>
<td>.00</td>
<td>.51</td>
</tr>
<tr>
<td>Experiment * age-group</td>
<td>13.31</td>
<td>2</td>
<td>6.66</td>
<td>15.88</td>
<td>.00</td>
<td>.40</td>
</tr>
<tr>
<td>Error</td>
<td>19.70</td>
<td>49</td>
<td>.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>88.30</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The general purpose of this study was to investigate the experiences of young children toward the smart storytelling toy StoryTech. Research questions of the study were as follows:
1. How does playing with StoryTech impact the children’s narrative activities?
2. How does playing with StoryTech impact the children’s creativity?

To answer research question 1, the discourse patterns of the children were analyzed. According to the findings, the 5- and 6-year old children who played with StoryTech, both alone and in dyads, talked more than the children who played with the passive toy. This is similar to Cassell and Ryokai’s (2001) finding that children on StoryMat, both alone and in dyads, paused less than the children on the passive mat. Although Cassell and Ryokai (2001) did not analyze the results in terms of age differences, this study provided findings for the 4-, 5-, and 6-year-old children. This study also provided such a large-scale user study with children including different age-groups. Based on the findings, children in the age 4 group who played with the StoryTech, both alone and in dyads, did not show a difference in terms of talking activity compared with the children who played with the passive toy. These age-group children also produced less stories than the children in the age 5 and age 6 groups. The reason for this situation, on the one hand, can be related to the view that younger children always need direct, clear, and immediate feedback to stay interested to the activity (Fontijn & Mendels, 2005). On the other hand, the 4-year-old children, both alone and in dyads, played with StoryTech more than the passive toy. As findings indicated, children who played with StoryTech alone in the age 5 and age 6 groups focused on storytelling more than the children who played with StoryTech in dyads in the age 5 and age 6 groups. This finding reaffirms that children in dyad sessions generally emphasize on play acting more than storytelling (Kehoe et al., 2004). The findings also indicate that children who played with StoryTech in the age 4 group, both alone and in dyads, received more support during their discourse process than the children who played with StoryTech in the age 5 and age 6 groups. Another interesting finding is that while the discourse time of children who played with the passive toy, both alone and in dyads, in the age 6 group was higher than that of the children who played with the passive toy, both alone and in dyads, in the age 4 and age 5 groups, the discourse time of children who played with StoryTech, both alone and in dyads, in the age 6 group was nearly equal to children who played with StoryTech, both alone and in dyads, in the age 5 group.

To answer research question 2, the number of imaginative objects was analyzed. According to the results, the 5- and 6-year-old children who played with StoryTech, both alone and in dyads, produced more imaginative objects than the children who played with the passive toy. Similarly, Cassell and Ryokai (2001) found that the children who played on StoryMat with or without a playmate transformed a higher number of imaginative objects than the children who played on the passive toy. Cassell and Ryokai (2001) also found that “the children who played on StoryMat alone transformed the highest number of imaginative objects” (p. 183). In our study, we extended this finding in a way that the children who played with StoryTech alone in the age 6 group produced the highest number of imaginative objects. Although, there was a considerable difference between children who played with StoryTech alone in the age 6 group and children who played with StoryTech in dyads in the age 6 group in terms of number of imaginative objects, this difference was not considerable between children who played with StoryTech alone in the age 5 group and children who played with StoryTech in dyads in the age 5 group. Based on the findings, there was no considerable difference between 4-year-old children who played with StoryTech, both alone and in dyads, and their counterparts who played with the passive toy, both alone and in dyads. The two-way ANOVA analysis results reaffirm that there is no main effect on the conditions children played in (single or dyads) and the main effect for the type of the toy (smart toy or passive toy) (Cassell & Ryokai, 2001). Based on the results of another two-way ANOVA analysis, the mean number of imaginative objects transformed by the age 6 group children was significantly higher than the mean number of imaginative objects transformed by the age 4 and age 5 groups.
Conclusion

According to Rogers et al. (2002), we need to examine systematically theoretical and empirical aspects of mixed realities to understand their potentials on different user experiences different from other computer-mediated environments. With the storytelling smart toy, the researchers presented a mixed-reality environment where children play with tangible toys and see an interactive environment on the computer. The findings have shown that such an innovative technology enhances educational power of traditional storytelling approach by combining tangible and virtual components. It stimulates cognitive processes and causes rich narrative activities. Results of the study will be helpful for educators who are interested in children’s narrative development via technology without completely separating children from reality. One of the most important properties of the toy is that both backgrounds and the characters can be expanded with new toys and objects whenever desired. This is important for the toy to be used in varied contexts and by children in different age levels. This promotes contextual and sustainable learning by using imagination. The simplicity of the toy encourages very young children to play with it individually or under teachers’ supervision and tell the stories. Results of the study also provide some implications about how we can moderate virtual and real parts of the toy to provide the best experience for children of different ages.

Acknowledgment

We would like to thank Turkan Karakus and Ismail Yildiz for their help in the development process of the StoryTech.

References


Kehoe, C., Cassell, J., Goldman, S., Dai, J., Gouldstone, I., MacLeod, S., et al. (2004). Out of the lab and into the world: Bringing story listening systems to the classroom. AERA.


Integrating Traditional Learning and Games on Large Displays: An Experimental Study

Carmelo Ardito, Rosa Lanzilotti, Maria F. Costabile and Giuseppe Desolda
Dipartimento di Informatica, Università di Bari Aldo Moro, Italy // carmelo.ardito@uniba.it // rosa.lanzilotti@uniba.it // maria.costabile@uniba.it // giuseppe.desolda@uniba.it

ABSTRACT
Current information and communication technology (ICT) has the potential to bring further changes to education. New learning techniques must be identified to take advantage of recent technological tools, such as smartphones, multimodal interfaces, multi-touch displays, etc. Game-based techniques that capitalize on ICT have proven to be very valuable. This paper presents a study aimed at validating a new educational format, inspired by the Discovery Learning technique, which integrates educational games, designed to be played on large multi-touch displays, with other types of formal and informal learning. Six classes of a primary school were involved in the study. The results showed that the proposed educational format is effective and that games on these novel multi-touch systems engage pupils very much, stimulate their collaboration and help consolidating knowledge.

Keywords
Large multi-touch display, Educational game, Formal and informal learning, Discovery learning, Field study

Introduction
The availability of new technologies, ranging from last generation mobile devices to gestural input, multimodal interfaces, multi-touch displays, etc., is bringing about changes in the field of education. Schools are under increasing political pressure to demonstrate that technologies actually improve student learning but assessments conducted on technology programs have not always yielded significantly improved student performances (Oviatt 2012). Our experience suggests that technological tools cannot replace the teacher, but they can certainly be considered a valid support for students in their learning paths (Costabile et al., 2008; Lanzilotti and Roselli, 2007). It is indispensable to define learning techniques that allow teachers to exploit technology to support students’ acquisition of new knowledge. Games based on information and communication technologies have proven to provide a valid support for effective learning (Cabrera et al., 2005; Lanzilotti and Roselli, 2007; Rogers et al. 2005). An educational game creates a pleasant learning environment, in which students learn more easily and enjoyably. Games foster relational skills, encouraging young people to work in groups and collaborate to attain given objectives. Each student carries out the activities s/he feels most congenial and, by working together, they can solve clues or problems, overcoming difficulties thanks to their common efforts. Innovative technologies, like mobile devices, multimodal interfaces, multi-touch displays, etc., can make the game more engaging (Oviatt 2012).

This paper describes a study that shows how more traditional learning carried out at school may be integrated with educational games that employ advanced technology, namely large multi-touch displays. To perform this integration, a new educational format is proposed, which is inspired by the Discovery Learning technique defined by Bruner in his Constructivism theory (Bruner, 1990). Bruner defines learning as an active process in which new information goes through three different types of representation: symbolic (based on language), active (based on action), and iconic (based on images). The educational format indicates how to provide pupils information by integrating formal learning activities, as performed in traditional school settings, with more informal and technology-based learning activities, in order to facilitate knowledge acquisition and also consolidating it.

The study was conducted to validate the educational format and evaluate the use of the multi-touch technology in school. It enrolled 107 pupils of six fifth grade classes at a primary school in Bari, Italy. Results showed that the educational format was effective and all pupils enjoyed the experience. The system implementing the educational game proved to be very engaging, able to stimulate pupils to work in group and to favor inclusion. Inclusion refers to the involvement in the school activities of all students, regardless their possible disabilities, as those derived from the social, cultural and personal context of the student (Ianes, 2012). Indeed, even those pupils, who were usually timid in class, were very much involved and actively collaborated with their companions.
The paper is organized as follows. The next section reports related works on multi-touch technology applied to educational activities. The third section presents the educational format. The educational game implemented on a large multi-touch display is described in the fourth section. Next, the field study and its results are reported. The paper concludes by discussing our findings and proposing directions for future research.

Related work

There is growing interest in investigating the use of multi-touch displays in different domains, thanks to their decreasing cost and increasing availability. Multi-touch displays have been used for gaming and entertainment (Tse et al., 2011), to support tourists (Jacucci et al., 2010; Marshall et al., 2011), etc. Researchers are now considering this new technology for educational purposes. We describe some examples in the following. Later in this section, we report three evaluation studies comparing the use of multi-touch displays with traditional paper and pencil and/or classic desktop applications.

One of the first examples of multi-touch technology applied to education was Read-it, a game-based application, designed to support the development of reading skills in children aged 5-7 years old (Sluis et al., 2004). The results of a pilot experiment, involving 15 pupils, clearly showed that pupils enjoyed playing the game and that the technology was not an obstacle to learning.

A few years later, Piper et al. presented SIDES (Shared Interfaces to Develop Effective Social Skills), a multi-player game that encouraged the use of cooperative skills by students in social group therapy (Piper et al., 2006). The results of two studies confirmed the authors’ belief that cooperative games are an effective paradigm for teaching effective group work skills, and that tabletop technology provides promising tools facilitating cooperative gaming experiences for special needs populations.

Antle et al. developed an application, called Futura (The Sustainable Futures Game), whose aim is to explore the novel design space for multi-touch tabletop games (Antle, 2011). The learning goal was to help people improve their understanding of the sustainable development. Futura allowed players to see how and what their co-players were doing, providing opportunities for each player to learn from and help others. A study performed with hundreds of users of all ages during the 2010 Winter Olympics indicated that Futura was effective and enjoyable. Children and teenagers seemed to have more fun than adults, who seemed intimidated by their first round of play. No players had any difficulty understanding how to interact with the table.

Fu et al. proposed multi-touch technology to deliver an effective interface to navigate the unique features of 3D environments such as astrophysical simulations (Fu et al., 2010). To investigate people's behavior with this new educational approach, the authors performed a study with 16 participants, who liked very much the physical and intuitive interaction with a large display just using their hands.

To the best of our knowledge, the only example of multi-touch display set up vertically, as in our case, is “Us Hunters” a multi-user application that allows children, during their visit a museum, to experience and learn about the hunting strategies, which are depicted in a painting on the wall of a cave underneath the museum, dating 6000 B.C. (Kourakis and Pares, 2010). No formal evaluation had been yet performed. The children experience was observed by a guide expert in prehistoric art, who reported that, on the basis of comments and questions children asked after playing with Us Hunters, the application had great potential to be an effective tool in the learning process. A formal evaluation study was planned as future work.

Some researchers have compared learning through multi-touch technology with learning using a paper and pencil or a desktop. Piper and Hollan, for example, compared the results of presenting educational material on a tabletop display with presenting the same material using traditional paper handouts in a neuroscience class of 20 undergraduates at the University of California (Piper and Hollan, 2009). Results revealed that students studying the materials on the tabletop were able to overcome problems on their own and repeated activities more often than students studying with paper material. Moreover, greater playfulness and enjoyment were noticed.

Soro et al. compared observations of user behavior in pair programming tasks performed at a traditional desktop versus a multi-touch tabletop, studying 44 students of Computer Science or ICT professionals (Soro et al., 2011).
Participants were asked to review 7 snippets through a simple interface, implemented with an identical look and feel for the desktop and the multi-touch table. Results showed that people performed better working at the multi-touch table, since it encouraged cooperation and helped people express their potential.

The study by AIAgha et al. aimed at understanding possible benefits of integrating multi-touch tabletop into classroom activities (AIAgha et al. 2010). They developed a system, called TablePortal, which allowed teachers to manage and monitor collaborative learning on students’ multi-touch tables. The teacher used a separate table to communicate with the students’ tables; in this way, teachers and students could work together on their multi-touch tables and collaborate on learning tasks. Observations in a real context showed an enhanced level of teacher’s awareness, flexible monitoring, and a positive impact on social interactions in the classroom.

Our study, too, is focused on the integration of large multi-touch displays in school activities by proposing a format that includes other educational activities, as illustrated in the next section.

The educational format

The integration of formal and informal learning proposed in this paper is inspired by the Discovery Learning technique defined by Jerome Bruner in his pedagogical theory of Constructivism (Bruner, 1990). This theory supports the belief that learners construct new ideas or concepts based upon existing knowledge. Thus, the process of learning is active and involves transformation of information through three different types of representation, namely symbolic, active, and iconic, based on language, action, and images, respectively.

In the symbolic representation, information is mostly in the form of words, mathematical symbols and other symbolic systems (e.g. musical symbols). In the active representation, pupils learn by performing physical tasks, e.g. manipulating objects, working in the laboratory, visiting places like museums, monuments, etc. In the iconic representation, visual images are used to illustrate the concepts to be learned. More in general, an iconic representation can be visual, auditory, olfactory, or tactile.

According to Bruner, all three types of information representation are essential and important. Learners activate, perhaps unconsciously, one type of representation or another, and sometimes switch from one to another. For example, pupils learn to ride a bike not only by watching someone riding or reading a book of instructions, but they have to be active themselves. Bruner emphasizes the “discovery” in the learning process, because it stimulates learners to “construct” their knowledge and to detect correlations and regularity among the information they have acquired in one of the three types of representation, helping learners to identify possible transformation to be operated on the information in order to better understand it (Bruner, 1990).

![Figure 1. Activities in the three phases of the educational format](image)

The educational format we propose integrates formal learning (traditional classroom lessons) with more informal and technology-based learning, organizing learning activities in three phases (Figure 1). Pupils get new information by: 1) attending the lesson(s) by their teacher in the classroom (symbolic representation), 2) acting in a real context (active representation), and 3) interacting with technological tools to manipulate images (iconic representation).
While the activity carried out in the symbolic phase is always the classroom lesson(s), those performed in the active and iconic phases change according to the topic to be learned and the technology used. For example, if the topic is natural science, in particular the naturalistic park habitat, pupils first attend lessons at school, then they explore the naturalistic park and lastly they interact with an educational system that allows them to manipulate concepts represented by images (e.g. a tree, a plant, etc.) showed on a digital screen.

In this paper, the educational format has been used to foster history learning and a keen interest in cultural heritage in primary school children aged 9-10 years old. Specifically, pupils learn about ancient Roman history by attending a lesson, in which their teacher provides basic notions on this topic (symbolic phase). After this, pupils learn during a school trip to an archaeological park of an ancient Roman city, or visiting a museum in which they watch how people used objects in the past. During the visit, the guide stimulates children to explore the site in order to discover signs of places and monuments of the ancient city. For example, the guide suggests: “Go in that direction and look for the rest of a road paved with stone blocks. You may see signs left by the wheels of heavy carts”. Finally, back at school, they interact with an application implemented on a large multi-touch display. This application, called History-Puzzle, is a game which proposes puzzles that pupils have to reassemble. The puzzles depict places or objects that pupils saw during the school trip.

A game on a large multi-touch display

The History-Puzzle game, considered in the third phase of the educational format, is implemented on a system with a large multi-touch display (Ardito et al., 2010). Such novel devices foster collaboration by allowing people to interact together at the same time. The display was set up in a vertical position to make it visible to other people standing nearby those currently interacting. We used a MultiTouch Ltd (http://multitouch.fi) 46-inch Full HD LCD display called cell. By mixing Rear Diffused Illumination and Led Light Plane technologies, this device is able to detect unlimited touch points. More cells can be linked to create greater screens. In our case, we used only one cell, not only for budget constraints but also because it was easier to transport and install. This device is very efficient in recognizing gestures, so even novice users can easily interact. Object manipulation is mediated by gestures performed with one or both hands (or some fingers). The main gestures are: Move, to move an object on the screen, which can be done with one hand just by touching the video; Scale, to change the size of an object, done with the index fingers of both hands, by stretching or narrowing two corners of the object; Rotate, to rotate an object, again using the index fingers of both hands. History-Puzzle has been developed in Java, with the open source MT4J (Multitouch for Java) development platform (http://www.mt4j.org). The TUIO protocol (http://www.tuio.org) has been adopted to allow communication between the touch-capture system and the software application.

Figure 2. The screen to start solving the puzzle “Via Traiana”

History-Puzzle is designed to be played by children interacting with a multi-touch display installed in a school laboratory, in an indoor space of an archaeological park, or in a museum. It is named History-Puzzle because groups of pupils collaborate to reassemble puzzles to discover elements of interest at an historical site. In the game played after a visit to the archaeological park of Egnathia, in Southern Italy, the puzzle images are significant places in this
ancient Roman city, e.g. the Civil Basilica (the Law Court), the Trajan Way (the main road from Rome traversing Egnathia), the Foro Boario (the animal market square), the Kiln (oven for cooking terracotta vases). For example, after a player selects a puzzle, such as the Trajan Way (Via Traiana in Italian), a screen like the one shown in Figure 2 appears. The figure to be discovered by solving the puzzle is at the center of the screen. The nine square tiles covering the figure contain incomplete sentences reporting historical notions about the selected place. For each tile, the players choose the rest of the sentence from the tiles displayed on the left and right sides of the puzzle and drag it onto the tile in the figure.

In the example in Figure 2, the sentence “Collegava Benevento a …” (“It connected Benevento to…” in English) is associated to the town name “… Brindisi”. If the selected association is correct, the tile will reveal one ninth of the 3D reconstruction of the original place. Figure 3 shows what it looks like when the player has discovered 7/9 of the image.

Some tricks are inserted to make the game more challenging and intriguing: 1) more than nine tiles are shown outside the puzzle: these additional tiles report false answers or answers that do not match any of the nine incomplete sentences displayed; 2) the positions of the tiles with incomplete sentences and those with completing sentences are randomly reassigned each time the puzzle is run; 3) sentences are randomly chosen from a file containing about 30 sentences for each puzzle. When the nine sentences have been completed and the whole image is displayed, a new game screen is shown. It proposes different multimedia contents: a papyrus reporting all the puzzle sentences, a short video showing the 3D reconstruction of the place, photos of the place, the half-section plan of a building (Figure 4).
The system also reproduces sounds associated to the place, various noises of typical activities carried out when the civilization of Egnathia was alive. In the case of the Trajan Way, carts and horses running are heard. Finally, the system returns to a screen showing the map of the park (Figure 5). The number “44”, displayed on the Trajan Way, indicates how many points the group gained for solving that puzzle. This score is computed by summing the number of correct associations and subtracting the number of wrong ones. A score of 5 points is awarded if the sentence is correctly associated, while the score is reduced of 2 points every time children move a tile on a wrong one. This score mechanism stimulates children to reflect upon their actions and leads them to discuss together on the tiles they have to associate.

![Figure 5. Map of the archaeological park showing the available puzzles](image)

**Field study**

A field study conducted during November-December 2011 at a primary school had three goals: 1) to investigate the pupils’ experience in the interaction with multi-touch displays in a real context; 2) to evaluate the effect of multi-touch technology as a means of consolidating knowledge; 3) to analyze the effectiveness of the proposed educational format.

**Participants and Design**

The study involved six fifth grade classes at the primary school “Clementina Perone” in Bari, Italy. A total of 107 pupils (55 girls, average age 10) participated in the evaluation as part of their school activities. For goal 2), a subset of 53 pupils (hereafter denominated the *multi-touch group*) belonging to three classes answered a true/false test twice: immediately before and after interacting with History-Puzzle. For goal 3), all pupils answered the same test only at the end of the educational experience.

**Procedure**

The study lasted four weeks and consisted of three phases: *symbolic*, *active* and *iconic*. During the first phase, participants attended a 2 hour lesson in which the class teacher presented basic notions about ancient Roman history focusing on Egnathia, a city which was active from the IV century BC. In the second week, the active phase was carried out through a school trip to Egnathia, in which they visited the archaeological park of Egnathia and the museum; a professional guide accompanied them in the park, helping pupils to identify meaningful places of the ancient Roman city, and later in the museum, where pupils saw how the ancient people used objects in their daily life.

During the third week, in the iconic phase, the six classes took turns, in groups of 4/5 pupils from the same class, to go to the school laboratory where the multi-touch display was installed, and play with History-Puzzle. All the other
pupils stayed in class doing other activities unrelated to Roman history. This was because the technical and human resources were sufficient to observe only one group playing the game at a time. Each group had a total of 15 minutes to play with the multi-touch display and solve three puzzles; they could freely decide which strategies to adopt to reassemble puzzles. After the whole class had interacted with the display, they answered a true/false test in the classroom to measure the acquired knowledge. Then, the procedure was repeated with pupils of another class. Three classes were observed on one day and the other three classes the day after; the multi-touch group answered the true/false test both before and after playing with History-Puzzle.

Pupils were observed by two research assistants, who noted down the main events and provided help during the interaction when explicitly requested. Two cameras were used for videotaping. Camera 1 was installed on a tripod placed to one side about two meters away from the system to film hand movements on the screen. Camera 2 was placed on top of the display to film the children's faces while they interacted with the system.

During the fourth week, a focus group was held in each class, moderated by the teachers. This served to gather information on the overall experience, starting from the lesson about Egnathia, passing through the school trip and finishing with the game on the multi-touch display. The six focus groups were video-taped and a research assistant noted down the pupils' most important observations.

Two months after the study, interviews to the four teachers of the classes involved in the study were performed. The aim was to get their views on the educational format by also considering the possible pupils reactions during the period following the study. The interviews were registered.

**Results**

The value of the proposed educational format, which integrates informal learning based on the use of a digital game, does not lie only in assessing how much pupils learned, but how they learned, i.e. how much they were engaged in what they were doing, if they liked it, if they enjoyed collaborating. Thus, in our study, we collected and analyzed a considerable amount of qualitative data. Quantitative data were gathered through true/false tests that have been administered to address learning; they were complemented with comments provided by the teachers during individual interviews.

In this section, we first illustrate results about learning then, we report the data gathered during observations of the pupils interacting with the History-Puzzle application, and their comments during the focus groups, in order to evaluate the overall pupils’ experience. Such data are presented in three parts, addressing pupils’ behavior, their performance and their own perception, respectively.

Three researchers analyzed a total of 5 hours, 46 minutes and 11 seconds of videos of 22 groups interacting with History-Puzzle. Researchers independently double-checked some 50% of the material. If the inter-rater agreement was less than 70%, the researchers discussed the differences and reached a mutual agreement. Reliability was high (agreement over 85%) for all the variables reported in this paper.

**Learning**

An initial quantitative analysis on the test administered to the pupils was performed to measure the knowledge acquired. The true/false test included 27 questions addressing different aspects of everyday life in Egnathia, illustrated by teachers during the lessons, repeated by the guide during the trip, and seen again on the multi-touch display. This test was administered to all pupils after the use of History-Puzzle to measure the effectiveness of the educational format. The multi-touch group did this test twice (before and after using the multi-touch display). A learning score was obtained for each assessment phase, summing the number of correct answers at each test and converting these results to decimals. On average, all participants obtained a learning score of 7.12 (SD = 1.02). The two learning scores of the multi-touch group were analyzed by t-test, that revealed a significant effect (t(99) = 2.83 p < .005). On average, participants answered 17.2 items correctly prior to the multi-touch interaction and 19.24 items after the interaction.
In the interviews performed to the four teachers, it emerged that, in their opinions, the educational format facilitated pupils to acquire contents that were presented in the different phases. Teachers said that there is evidence that the knowledge gained from the game remained very much impressed. They noticed that, during school activities performed some weeks after the study, in which there was the occasion to discuss concepts of the ancient Roman life, pupils often made reference to their experience with History-Puzzle and with the visit to the archaeological park. For example, during a visit to the ancient area of Bari, a month after playing the game, looking at the rest of the Trajan Way in Bari, some pupils explained: “Oh, it is the same road we saw at Egnathia and on the multi-touch display!” Indeed, teachers’ experience shows that children of this age learn a lot by observing images. In addition, the use of a multi-touch display involved children at sensory level by allowing them to touch with their hands images and to interact with them.

All teachers highlighted that pupils felt protagonists, especially during the visit and the game. A major positive result was about pupils’ inclusion (Ianes, 2012): those pupils, who in class seem apathetic, not motivated and tend to distract, actively participated in the game, provided appropriate answers and contributed to the group activities with enthusiasm.

Teachers also noticed that, when children interacted with the multi-touch display, they made spontaneous reflections and self-assessments. In particular, three teachers reported that they often heard the children explaining very proudly “Yes, I answered right here!” or saying sadly: “No, the answer I gave was wrong!” Another important aspect that teachers reported was collaboration. Generally, in laboratory activities, children work in pairs and sometimes individually. With the multi-touch display, children enjoyed work in larger groups and interacting with their peers.

**Behavior**

In total, the 22 groups employed 3 hours, 58 minutes and 19 seconds to solve the three puzzles; on average, a group spent 10 minutes and 50 seconds to reassemble the three puzzles. The remaining time was used by participants to: decide which puzzle to solve, watch the multimedia contents available on the display after each puzzle, decide on the name their group and see their scores.

During the study it was observed that multi-touch display stimulates collaboration encouraging group activities. In fact, participants generally collaborated a lot to reassemble the puzzles. Only in 8 out of the 22 groups did each pupil in a group work alone to compose parts of a puzzle. Different strategies were used: a) some groups divided the tiles according to the members’ position in front of the display, as illustrated by the following excerpts: “You do that part because you’re on the right and we’ll do this one”; b) other groups established impartial rules: “Let’s take turns, one tile each”; c) other groups decided that the puzzle should be built by each member who proposed a possible right answer: “You said it so you do it”.

The groups mainly stayed united and collaborated to reassemble the puzzle. If one member tended to draw back the others encouraged her/him to take a more active part. This cohesion was particularly evident when they got into difficulties, as demonstrated by the following pupil’s exclamation: “United we stand” (see Figure 6).
In 13 out of 22 groups, there was no dispute (Table 1). In 5 groups the pupils disagreed only once, in 2 groups twice, 4 times in one group, and 6 times in another group. This last group quarreled about the use of the display because one member wanted to hog the whole game.

<table>
<thead>
<tr>
<th>Dispute situations</th>
<th>N. of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>13</td>
</tr>
<tr>
<td>Once</td>
<td>5</td>
</tr>
<tr>
<td>Twice</td>
<td>2</td>
</tr>
<tr>
<td>4 times</td>
<td>1</td>
</tr>
<tr>
<td>6 times</td>
<td>1</td>
</tr>
</tbody>
</table>

Analysis of the group dynamics during the game addressed the leader role, which is negative in this case. In fact, a leader role means that one pupil tends to impose her/his will on the others, dictating which tile should be moved and why. This was often accompanied by dominant gestures like blocking another pupil's hands to prevent her/him from touching the tiles. Only in 36% of the sample did a pupil attempt to play a leader role. These members were not well accepted by the rest of the group, who all wanted to take an active part in solving the puzzle, and rebellion and confusion ensued: “Michele, stop, you've already done two! Let us play too!”

Competition is an aspect that often emerges in gameplay. A reasonable degree of competition is considered positive to motivate and stimulate players to do their best. In our study, intra-group competition did not really occur because the pupils in the group mainly collaborated to carry out the game. Each group appeared interested in doing better than the other groups and they all checked their final scores against those of the other groups on the overall classification the system shows on request. Moreover, in 11 out of the 22 groups there was at least one explicit competition episode (Table 2), that emerged when one or more members of the group insisted that they needed to get as high a score as possible (e.g., a child said: “Let's try not to make a mistake so as to get as many points as possible”).

<table>
<thead>
<tr>
<th>N. of competition events</th>
<th>N. of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

The multimedia contents available at the end of each puzzle composition were generally not found interesting because the groups wanted to go on to solve the next puzzle. Table 3 shows the time each group spent watching such contents. It was assumed that 15 groups were not very interested because they observed the contents for only 1-4 seconds. The 5 groups that spent 5-10 seconds watched the initial part of the video and some photos, whereas the last 3 groups actively interacted with this multimedia content, and had fun zooming in and out, and rotating it.

<table>
<thead>
<tr>
<th>Time (seconds)</th>
<th>N. of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>15</td>
</tr>
<tr>
<td>5-10</td>
<td>4</td>
</tr>
<tr>
<td>11-22</td>
<td>3</td>
</tr>
</tbody>
</table>

A lot of pupils enjoyed doing the puzzle; this could also be seen by the fact that they accompanied the movement of the tile with their whole bodies: they knelt while they moved the tile down or leaned over to the right or left as they moved the tile in that direction.

**Performance**

The first puzzle was used to understand game and interaction mechanisms, and to get organized, and then they were able to solve the next puzzles more rapidly (Table 4). Although the three puzzles had the same difficulty, pupils
employed about 29% of the total time to reassemble the first puzzle, decreasing to 24% for the second puzzle and only 17% of the time for the last puzzle.

Table 4: Percentage of total time pupils employed to perform game activities

<table>
<thead>
<tr>
<th>Game activity</th>
<th>Time (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solving the first puzzle</td>
<td>29</td>
</tr>
<tr>
<td>Solving the second puzzle</td>
<td>24</td>
</tr>
<tr>
<td>Solving the third puzzle</td>
<td>17</td>
</tr>
<tr>
<td>Observing multimedia contents about puzzles</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>25</td>
</tr>
</tbody>
</table>

Pupils had no problems in understanding the game mechanisms, but they encountered some problems of a technological nature (Table 5). In fact, the screen was less sensitive at the margins and in a central strip about two centimetres wide, making interaction with the object more difficult at these points. In these cases, too, pupils collaborated to solve the technological problem: if one of them had understood how to overcome the problem, she/he suggested the solution: “Drag it upward otherwise it thinks we want to put it there and won't go, indicating an error”. Unfortunately, only 2 out of the 22 groups did not encounter a technological problem but 12 of these groups had only one problem that they succeeded in solving alone (Table 5).

At the beginning of their interaction, if children had to associate a side tile with a puzzle tile located on the opposite side of the screen, their first attempt was to move the tile across the puzzle area. This was interpreted by the system as the association of the tile with the first one that was overlapped, generating a mistake. However, children soon understood that they had to reach the right tile by going around the puzzle area.

Table 5: Number of technological problems arising with the game

<table>
<thead>
<tr>
<th>N. of problems</th>
<th>N. of groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3-5</td>
<td>4</td>
</tr>
</tbody>
</table>

Perception

In previous studies, the questionnaire was used to gather data on pupils’ perception of the experience [8]. We found out that pupils were not able to clearly express their opinions. Thus, for this study we decided to perform a focus group in class to collect pupils’ comments on the overall experience.

The focus group data showed that participants tended to share knowledge acquired during the game, as well as about how to interact with the new technology. The participants explicitly declared that they greatly enjoyed the group experience they had with History-Puzzle, because it allowed each of them to remember, reflect and share knowledge with their peers. They said that the sentences they had to associate during the game were very useful to remember concepts learnt during classroom lessons and to better understand some aspects they had missed during the visit because sometimes they were distracted. They pointed out that they would prefer even a larger display to permit more groups to play together, rather than having to take turns.

Among the three types of information representation, they particularly appreciated the visit to Egnathia, because it allowed them to see what they had learned with the teachers, and also to admire cultural and archaeological heritage located in their territory, stimulating their powers of imagination and observation. In fact, they said things like: “It's like really living in this place”, “I felt like an explorer!”, “I imagined the people of the time riding down the Trajan Way in carts”, “No, I imagined people working in the mills”.
Discussion

The study reported in this paper was focused on three main issues: 1) investigating pupils’ experience in the interaction with multi-touch displays in a real context, 2) evaluating the effect of multi-touch technology as a means of consolidating knowledge, 3) analyzing the effectiveness of the proposed educational format.

As to the first question, the study demonstrated that pupils enjoyed the game and were actively engaged; they liked playing in groups and would have liked to play longer and with more companions. After the first puzzle, pupils often decided to divide the tasks among themselves so as to solve the problems faster. Multi-touch technology is very promising in facilitating collaborative experiences, as also indicated in (Piper et al., 2006). In fact, collaboration among pupils was the most important aspect observed during the study. The groups tended to work together, helping others who got into difficulties, suggesting solutions to problems emerging during the play, and exchanging information among them. Even if in some cases one pupil tried to take over the game and become the leader, this annoyed the other players because they preferred to work together. The study also revealed minimal competition among groups that preferred to go on working on the puzzle rather than watching photos or reading information given them after each puzzle. The game was very engaging, as shown by the fact that children moved their whole bodies in time with movements of the tiles.

As regards the second question, the study highlighted the value of the multi-touch technology as an important means to support knowledge consolidation. Pupils gave more correct answers to the true/false test performed after the multi-touch, confirming the results reported in (Sluis et al., 2004) that technology is not an obstacle for pupils but, on the contrary, can be effective to support learning processes. This result has been reported also by the teachers that were interviewed. The major surprise for the teacher was to discover that also pupils that in class are less motivated, during the interaction with the multi-touch display actively participated providing appropriate answers.

Finally, the data gathered during the focus group, the true/false test and the teachers’ comments provided further evidence that the three types of information representation (symbolic, active and iconic) are essential and important, and each reinforces the others. The symbolic representation was useful to introduce the new concepts. The active representation stimulated children's imagination and observation, and engendered a strong interest in history by allowing them to see and hear what life was like in ancient times. Finally, the iconic representation using multi-touch technology was useful not only in engaging pupils’ interest and enthusiasm while reinforcing the knowledge acquired, but also in stimulating collaboration and group work. The learning triad was successful as shown by the final assessment tests revealing that, in general, pupils reached a good knowledge level.

Conclusion

The introduction of ICT in education has encouraged the expression of information in different representations, modalities, and linguistic codes to stimulate ideas, clarity of thought, and improved performance in educational activities (Oviatt, 2012). This is especially evident when recent technological tools, like smart phones, multi-touch displays, etc. are used, but new learning techniques to fully exploit all these technological potentialities need to be defined.

In this paper, we showed that educational games using multi-touch displays are able to promote pupil learning, especially when games are integrated with other informal learning, such as a school trip, as well as traditional school lessons. This integration is proposed through an educational format that refers to the Discovery Learning technique, defined by Bruner (Bruner, 1990). According to this format, pupils learn through three different phases: 1) attending the lesson(s) by their teacher in the classroom (symbolic phase), 2) acting in a real context (active phase), and, finally, 3) interacting with technological tools (iconic phase), like the educational game on a large multi-touch display described in this paper.

Results of the field study showed that pupils were actively engaged in all the educational activities, indicating that the proposed educational format is effective and that applications on the multi-touch display can be a valid means for consolidating knowledge.
Bruner’s theory does not indicate any particular sequence in using the different representations for providing information. In designing the educational format, we decided to implement the technology-based iconic representation in the third phase. This decision was taken after discussions with teachers, who prefer to use technological tools to deepen knowledge and practice new concepts which have previously been introduced in formal lessons. Future work will include further studies to investigate different sequencing of the three phases in the educational format.

Acknowledgments

This research was partially supported by Italian MIUR (grant "Learning4All"). We thank teachers and students of the school “Perone” in Bari, Italy, for participating in the field study. Special thanks go to Prof. Floriana Falcinelli of the University of Perugia, Italy, for her useful comments on pedagogical issues.

References


Affectibility in Educational Technologies: A Socio-Technical Perspective for Design

Elaine C. S. Hayashi and M. Cecilia C. Baranauskas

Institute of Computing & Nucleus of Informatics Applied to Education, University of Campinas, Av. Albert Einstein, 1251, Campinas, SP, Brazil // hayashi@ic.unicamp.br // cecilia@ic.unicamp.br

ABSTRACT

Digital artifacts have the potential for augmenting the interest of students and the quality of learning environments. However, it is still common to find technology being inserted in learning settings without a closer connection to the learners’ contemporary world. In this paper we report on results of a qualitative research conducted to address questions such as: How could a new technology be introduced in schools in a way that it makes sense to the users? Could it contribute to more integrated learning scenarios? The study took place at an elementary public school in the city of Campinas, in São Paulo, Brazil, and involved more than 500 people. Data were collected from participatory Workshops, informal interviews and pictographic questionnaire. The results suggest that it is possible to combine school’s formal and informal practices into meaningful learning by using the XO educational laptop. The activities helped teachers and students realize that the use of technology can be recreated, influenced by their own feelings, values and culture. A discussion is conducted towards new perspectives on understanding learning practices mediated by technology and the role of the concept of ‘Affectibility’ in the design of educational systems.

Keywords

Human-Computer Interaction, Design for Children, Laptops in Education

Introduction

There seems to be a common sense – not always agreed – that formal learning refers to learning that takes place inside the school; and informal, outside the school (Eshach, 2007). Non-formal learning is often seen as learning that also takes place outside the school, like in informal learning, but still has the same “deliberate instructional and programmatic emphasis” as in formal learning (La Belle, 1982). Similarly, Looi et al. (2010) define formal learning as the one “which is based on fixed curricula enacted in classroom environments” (p.156). Coombs and Ahmed (apud La Belle, 1982) characterized formal learning as the “institutionalized, chronologically graded and hierarchically structured educational system, spanning lower primary school and the upper reaches of the university” (p.162). Still with the emphasis on the physical space where learning occurs, Looi et al. (2010) define informal learning as the type of learning system “where learners are participating in intentional or unintentional experiences outside school settings” (p.156); and non-formal learning, referring to “learning that happens in formal learning settings but not tested or assessed in traditional ways”.

Other aspects taken into consideration while differentiating formal, informal and non-formal learning include: the control learners have over the learning situation (Mocker & Spear, 1982); and the presence or not of a mediator and his role (Eshach, 2007). Another important aspect refers to the intentionality of learning: in formal and non-formal, the intention of learning seems to always be present, in varied degrees (Eraut, 2000; Schugurensky, 2000). In informal learning, the intention of learning may or may not be present (Schugurensky, 2000).

In spite of divergences that might be found among these definitions, “informal”, “formal” and “non-formal” seemed appropriate to describe learning practices in the traditional educational models so far.

The act of teaching (actually, the act of guiding learning processes) demands reasoning and comprehension of reality, in a dialogical approach. Paulo Freire, one of the most influential educators in Brazil, had always criticized the way educational approaches are often detached from learner’s real life contexts (Freire, 1968). As the digital culture is sculpturing a new reality to which school’s practices need to be adapted, this fact calls for a transformation in schools. In a dialogue about Education and Technology between Paulo Freire and Seymour Papert – the creator of LOGO language, who proposed the Constructionism learning theory (Papert, 1993) –, Freire stated that the purpose is not to discontinue the school, but to change it entirely, in a way that it can be reborn into a new, updated being (Freire & Papert, 1996).
In order for the school to catch up with its own time, it needs to incorporate technological advances in a critical and democratic manner. Even though technology is integrated in our lives, the ways it appears in education do not seem to be connected to the contemporary style of life. Moreover, while many children are being born in cultures where technology is naturally part of the environment, in Brazil, part of the population still has limited or no access to it. However, simply bringing a foreign technology inside the walls of a school might not solve the problem, as such technology might seem detached from the school’s reality.

Reviewing the meanings of formal and informal learning might help unraveling the new possibilities presented by digital technology, towards more seamless learning scenarios. Making sense of technology, in the context of children’s education, demands a socio-technical perspective that might contribute to the dialogical approach proposed by Freire (1968). Dialogue involves carefully listening to and understanding one another; it also includes the awareness of each other’s values, emotions and interests. In this paper, we explore the potential of the Affectibility concept (Hayashi & Baranauskas, 2011) in different scenarios throughout the use of low cost educational laptops, considered in the formal, informal and technical perspectives, as proposed by the Organizational Semiotics (Stamper, 1993; Stamper et al., 2000). Through some practical cases we address questions such as: how can a new technology be put into action in schools in a way that it makes sense to the users? Could that technology contribute to more integrated learning scenarios? We speculate on how the school’s practices (school’s values and culture) can point toward a meaningful use of technology.

The paper is organized as follows: in the next section we present the theoretical and methodological references; then we describe our context and research plan; present the results; discuss and conclude.

The compound layers of societal design

In today’s digital world, technology breaks barriers of space, time, and of presence or absence of mediators. It let learners and teachers experience out-of-the-school activities while still inside the classroom. While outside the school, learners may watch video or browse through and interact with a multitude of learning contents. With the availability of digital resources, neither physical place nor human mediation seems to mark the limits of learning opportunities. The opportunities for the construction of learning are present anywhere, anytime, as well as the mediators. Also the forms of control vary. For example, there are virtual tours where learners can choose the type of experience and slide presentations that allow learners to choose and go directly to the information needed. Once the divisions demanded by space, time, and presence or not of mediators are broken, then the differentiation of formal, informal and non-formal learning – which were based on those divisions – should be reviewed. As Chen et al. (2010) put it, learning can occur seamlessly across time and places when properly mediated by digital devices. When a digital artifact is used as mere alternative and as a detached element, its potential is wasted. As pointed out by Barab (apud Looi et al., 2010, p.156): “While it is not to say that abstract knowledge and media-replicated experiences are not desirable in learning processes, one of the critical problems in traditional schooling practices is the excessive amount of decontextualized information, indirect and abstract knowledge, and secondhand experiences confined in classroom contexts.”

For technological artifacts to serve its target community of users, as an embedded learning constituent, it is crucial to understand the way that the organization – - in this case, the entire school (Baranauskas, 2009b; Hayashi et al., 2012a) – works. The socio-technical approach enlarges the scope of learning settings, moving from perspectives based on place (i.e. inside and outside the school) to a perspective that considers informal, formal and technical aspects of the school context. We argue that this perspective is needed in the design of educational technology that aims at a seamless integration of technology into the contemporary school. In this approach, cultural aspects are taken into consideration, building a perspective from which the sense made by the use of technology plays a major role.

Cognitive models that used to inform the design of systems are undergoing transformation. As Boehner et al. (2007) suggest, researchers are enhancing the notion of cognition as more than rational thought. “Cognition has been proposed instead as something social and cultural, embedded in our everyday practices of making sense of and interpreting events.” (Boehner et al., 2005, p.60).
Culture is understood here as “the way of life of a people, the sum of their learned behavior patterns, attitudes and material things” (Hall, 1959, p.20) and it can be thought of as a “productive phenomenon” that “shapes individual and collective experience and gives it meaning” (Boehner et al., 2005, p.64). The environment of a community of users (e.g., a school, schools from a city or country) might change – or be changed by – the way technological device(s) are used. Since culture varies from people to people, as well as their language and the sense they make from signs, it is important to consider culture in the design and study of educational systems, as proposed by socio-technical approaches.

Cultural conventions (the way humans operate) can be classified in formal, informal and technical levels (Hall, 1959). The concepts of formal and informal from Hall differ from the definitions previously presented. Apart from semantic differences, the important distinction that can be stressed here is on how Hall’s categorization does not seem to break learning apart. Rather, it contributes to the desired integration; as put by Hall (1959, p.66): “One more generalization that should be kept in mind about formal, informal, and technical integrations is that while one will dominate, all three are present in any given situation.” Analyzing aspects in the three levels helps us be aware of the elements, and together these elements compose the entire organization.

Elaborating from Hall’s informal, formal and technical modes, the Semiotic (or Organizational) Onion (Stamper, 1993) helps us understand how each mode (or layer, referring to the onion metaphor) is embedded into the other, composing more meaningful information systems. Figure 1 illustrates and explains the layers of the Semiotic Onion from Stamper (1993) and how it works as a base for the science of design and participatory activities (Baranauskas, 2009).

In addition to the challenge of designing for digital learning scenarios, there is another factor to be considered. With the advances in technology, computer systems are no longer there for the sole purpose of helping us to complete our tasks. Devices are now part of our lives and are present everywhere. In this context, interaction design faces new challenges. As Abowd & Mynatt (2000) point out, usability techniques were appropriate when users’ tasks were known and system’s evaluation had the objective of verifying whether the interface was suitable for completing that task. Now “it is not at all clear how to apply task-centric evaluation techniques to informal everyday computing situations” (Abowd & Mynatt, 2000, p.51).

Within a socio-technical perspective and building upon less explored faces of usability, we have argued in (Hayashi & Baranauskas, 2011) that Affectibility should be taken into consideration in the context of educational technology. We proposed the concept of Affectibility as a guiding notion from which the design of system-user interaction could be treated. Affectibility refers to the aspects that make the system of good – or bad – affective, emotional and/or hedonic qualities, potentially evoking certain affective responses in the users.
Elements of the technical layer are often seen as deprived of emotion or affect (Hall, 1959), but the Semiotic Onion helps designers of technology to realize such aspects that might be left unnoticed otherwise. At the same time that affective issues can find its elements in each layer of the Onion, the most significant result is in the intermixture. By understanding the informal, formal and technical elements towards the design of technology, one may attain the understanding of the underlying base that is formed by the context of use where affective elements have a role to play.

Affectibility within the socio-technical scenario

Context

The OLPC (One Laptop Per Child) organization donated 520 laptops to a research Project (XO Project, 2010) being conducted at Padre Emílio Miotti School (Miotti): a public school located in the suburban area of the city of Campinas, in Brazil. The school has around 530 people (among teachers and students). Children’s age range from 6 to 14 years old and they are attending the fundamental level. They are all using the laptop, but it is not every grade that follows the one-to-one model (a few children share the laptops). A research project is being conducted within the school to investigate and construct a situated methodology for the laptops appropriation (Hayashi et al., 2012a).

Research plan and methodology

Our main research questions aim at understanding how to insert a technological artifact (the XO laptop), which was built in different context and culture, into the school practices of another country, preserving this country’s own cultural values. What are the main challenges that may emerge during this process? In order to make the technology’s use more meaningful, what are the aspects that should be considered?

![Figure 2. Framework for the study of Affectibility in interaction design of educational digital artifacts](image)

According to Miles and Huberman (1984), a conceptual framework maps the territory being investigated and this map should change over time. It specifies who and what will be studied and identifies relationships. Figure 2 illustrates the scope of this research as it depicts an improved version of our conceptual framework (a preliminary version was presented in Hayashi and Baranauskas, 2011). On the background of the figure lies the Semiotic Onion (Stamper et al., 2000). Figure 2 shows the involved parties in our research: users (school’s community), designers of
educational technology and the technology itself. The figure represents the macro view of our research. The micro
scope of our investigation, presented in this work, involves students and teachers (users – as the unit of analysis) and
their environment (school’s settings and other interested parties, who also took part in our activities).

All teachers, some other employees from the school (e.g. principal, pedagogue, janitor, cook, library attendant) and
the students participated in the activities that we describe in this paper. The data collection took place during
Workshops, where discussions and other activities were conducted. The Semio-participatory Workshops – SpW
(Baranauskas, 2009, 2009b) articulates principles and artifacts from Organizational Semiotics – OS (Stamper, 1993
and Stamper et al., 2000) and Participatory Design – PD (Muller, 1997) into practices that provide rich qualitative
data.

In order to improve the credibility of the data collected during the SpW’s, researchers have also been engaged in
activities at school, actively participating in regular classes for almost one year as Participant Observers (Gold,
1958). During that time, researchers interacted with students and informal interviews took place. Reliability
(audibility) was reinforced by pictures and video, together with researcher’s field notes. Like other results from
qualitative approaches – especially of Action Research (Avi son et al., 1999) -- transferability is recommended over
generalizability (Hayes, 2001). In this sense, the detailed description provided here might inspire other researchers to
apply similar approaches in their contexts, when inserting a new technology at a school.

During the researchers’ prolonged engagement within the school, all participants were aware of the presence of the
researchers and their objectives. The parents signed terms allowing the participation of their children in the research.
Teachers and other adult participants from the school also signed similar terms.

Results

In this section we present cases that illustrate some of the activities that took place at the school. The information
was gathered as results of the SpW and Participant Observations. The cases contribute to the understanding of how a
new technology can be inserted into more meaningful practices.

Case 1. Transforming homework assignments

The resources available at the laptop (e.g. web browser and wireless internet connection) allow students to promptly
perform tasks that would normally require a search in the library or a walk to the computer lab. Such visits are
essential, but the practicality of being able to perform some research in the classroom also has its advantages.

Cooper (1989) compared homework with in-class study and showed that in-class study results were superior.
Nonetheless, on another study, Cooper et al. (2006) identified evidences for the positive influence of activities
(performed at home) on students’ achievement. As pointed out by Xu (2005), activities sent to be done at home can
be an emotionally charged event, generally causing negative emotional states on children. That might be an
opportunity for parents and children to develop emotion management strategies (Xu, 2005). With their own laptops
available at hand, students are now capable of performing, in the classroom, some of the activities that used to be
assigned as homework allowing also teachers to take part in these emotional processes.

Activities that involve searching through sources that were not easily reachable when inside the classroom are now at
hand. Doing homework-like activities inside the classroom gives teachers the opportunity to understand some of the
emotional responses their pupils have during the discovery of the assigned topics. Furthermore, teachers are able to
help students in some of the difficulties faced when learners are doing research. Some of these difficulties, according
to Eilam (2001), are related to understanding what they are supposed to do and to making sense of the result of the
assignment. It is worth noting that we are not arguing for all homework to be transferred into classroom. As Cooper
et al. (2006) highlighted, homework has many benefits such as “greater parental appreciation of and involvement in
schooling”, “parental demonstrations of interest in child’s academic progress” (p.7), among others.

At Miotti, a teacher tried to bring to the classroom an activity that she would normally assign as homework. She
asked students to find out the origins and meanings of their names. If the search were made at home, students would
come with the answers ready and she would not have been able to take part in the process and directly observe the affective responses evoked by the discoveries they were making.

The teacher reported about a student who usually displayed low self-esteem behavior as he felt excluded for presenting some degree of learning disability. That student found out that his name (Victor) meant “victorious” or “he who conquers”. He also read about who Victor Hugo was. The teacher was excited when reporting to us about Victor’s attitude towards the result of his task: he was both proud and happy. Certainly, for the teacher, it was also a positive experience to take part on her student’s process of construction of self-image and identity. In this same practice, a girl found out that her name was originally a male name; the teacher took the opportunity to create a relation of trust by discussing this issue with that girl.

With the resources from the laptop, homework assignments – shown by Xu (2005) as an activity that usually evokes rather negative emotional responses – can now not only be brought as an inside-the-classroom activity but also can become something interesting and fun.

**Case 2. Integrating the School in Interdisciplinary Activities**

Other interesting activities resulted from a collaborative practice that started during the initial SpWs at the school. The objective was to collaboratively create Scenarios of Use for the laptops, in a way that it would make sense to students and teachers.

Scenarios are descriptions that allow us to reason about situations of use, even when the situations are not yet implemented (Carroll, 1999). These descriptions tell stories about people and their activities, including descriptions of places, events, actions, as well as objectives and changes that may happen throughout the stories (Carroll, 1999). As we have confirmed with this activity, scenarios might not only help designers to create systems but also helped in the anticipation of ways of embedding the new technology into their practices.

In order to create scenarios for the use of the laptop at Miotti, researchers, students, teachers and other employees from the school gathered together to propose a different scenario of use for the laptops; most propositions resulted in interdisciplinary activities. One of the resulting scenarios was called “Students and Consumption at Home”. In this scenario, the students used their laptops to take pictures of products (e.g., grocery, pharmacy, clothing) or product advertisements (from newspapers, flyers, magazines) to study different matters. Once the pictures were taken, further work on them would be performed, according to the subject. The activities that emerged from this scenario were related to different disciplines and some of them are described below.

The Science teacher asked students to take the laptop home and take pictures of the ‘nutrition facts’ labels for food products they most usually eat. From the information on the pictures, the teacher initiated a discussion about nutrition and later they built graphics (using the resources from the laptop) regarding food composition. They were able to compare calories and other information among products, constructing knowledge towards the promotion of healthier lifestyles.

From the same pictures of labels taken with the laptop’s webcam, the Portuguese (students’ mother language) teacher asked her teenager students to write diaries about their habits. The presence of the pictures of the labels of products in their diaries directed the content of the texts. This allowed students to be more aware of their consumerist habits. More details on this activity can be found in (Miranda et al., 2011a). For the Math class, the teacher asked students to take pictures from the electricity bills at their homes. From the analysis of those pictures, in the classroom, they studied about kilowatts and electricity consumption. Students also built graphics comparing the use of electricity along the months, and they discussed about ways of reducing the electricity bill. Students were supposed to share this information with their family in an effort to pay less for the bill and promote energy saving. Picture taking and merging them with texts would not have been possible to these students without the XO laptop, as many of them cannot afford digital cameras or cell phones that take pictures.

The English as a Second Language teacher asked them to take pictures from anything that was written in English, from labels in products to newspaper and magazine ads. This activity made students aware of the presence of the English language in their lives and also helped them to understand the meaning of those foreign words (as examples,
we can mention the words: “Dove”, “Ice Tea”, “reach”, etc., which appear either as trademarks or adjectives of/for commercial products).

These are only a few examples of the activities conducted at the school that were derived from that single scenario – out of the eight scenarios created (for more details on the other scenarios, refer to Amiel et al., 2011 and Miranda et al., 2011b). Most of the time, educational contents were discussed, in a collective creation of knowledge, without the students even noticing that.

Teachers frequently used the words “happiness” and “sense of accomplishment” to describe students’ attitudes during the activities. All teachers agreed that students seemed to be more interested and motivated to do the tasks using the laptop, either at home or in the classroom. Some teachers observed that students who used to present indiscipline problems were usually the ones to be more proactive with the use of technology. Instead of disrupting the class, these students, in the task of helping their fellows in the use of the laptop, were rather focused and committed. They were proud of their knowledge and aptitude with the use of the laptop. Furthermore, the interdisciplinary approach brought to learners a sense of engagement.

Case 3. XO inside and outside the school’s walls

All students seemed to be very happy with the use of the laptop in their daily activities, inside or outside the school. On the Independence Day, students went out to the traditional annual parade. The only difference this year was that a group of students were exhibiting their XO laptops. They were proud to show that they were having this opportunity of using the laptop.

In another experience, during a school trip, teachers reported that students were happy to explain about the laptop to bystanders who saw them using it. In this activity, students from the third grades visited a park in the city. Teachers asked them to register the animals they saw, as well as the information about the animals, either by taking notes, making short movies or taking pictures – using their laptops. The interesting point in this activity outside the school was that they were using the same laptop they use inside the school and, sometimes, at home. In similar studies reported in literature, where students experimented the use of technology in activities outside the school, they were usually given specific digital artifacts for each setting outside the school. For example, in Woods et al. (2004), the authors report on examples of educational applications (from different settings: museum, science exhibit, libraries, etc.) in which a different handheld device was necessary. As the device belonged to the place, students were not able to use the same artifact in different contexts (e.g., the handheld from the museum belonged to the museum and users were not allowed to take them to other places). That does not match the relationship that most people develop with technology in the contemporary world: in the same single device, which is to be taken everywhere, one can find their agenda, alarm clock, camera, digital watch, cell phone, word processor, web browser, etc. At Miotti, the children, working with their own XO laptop both inside and outside school, were able to develop values like ownership, responsibility and autonomy.

![Boxplots comparing the results obtained from the three Groups for the three dimensions](image)

*Figure 3. Boxplots comparing the results obtained from the three Groups for the three dimensions (SAM’s pictographic scales are represented in the Y axis. The higher the responses appear in the Y axis, the more positive they are. Refer to (Hayashi et al., 2012b) for more details.*)
The demonstrations of happiness and pride mentioned earlier were not only identified by observing students’ actions and expressions during the field observations (Participant Observation) from our qualitative research approach. We have also assessed their affective response to the XO laptop in a quantitative way. For that we applied SAM, the Self-Assessment Manikin (Bradley & Lang, 1994): a pictographic evaluation tool that captures the affective states of the subjects regarding an object or a situation. SAM has been successfully used in researches from different areas, as shown in (Hayashi et al., 2012b). Some results from initial assessments are reported in (Hayashi et al., 2012b) and illustrated in Figure 3. The results confirm the positive responses observed from the students. SAM assesses emotional responses in three dimensions: valence (or pleasure – indicating the positive or negative emotional reactions); arousal (or excitement or activation – indicating the bodily activation from the experience); and dominance (describing the feeling of being controlled or being in control of a situation or object). Figure 3 shows the results from three groups: younger children – from 6 to 10 years old (Group A); older children – from 11 to 14 years old (Group B); and teachers and employees (Group C). The result indicates that, in general, younger children present higher responses than older children (valence and arousal), even though they are not completely sure about how to use it (dominance).

Case 4. Student Volunteers

When the use of the laptops started to increase at Miotti, a new challenge was faced. The technical support team was not enough to respond to all demand from the more than 500 students plus teachers. The school did not have financial resource to hire people for the job. In order to cope with that problem, the school tried to follow the example of other schools that had similar laptop programs and encouraged students to volunteer as assistants. Similar to lab monitors, these students would help the teachers and their peers by giving them technical support for the use of the laptops.

Colvin (2007) calls this kind of activity Peer Tutoring as it “involves those of the same societal group or social standing educating one another when one peer has more expertise or knowledge” (p.166). In the project from Schultes et al. (2004), the responsibilities assigned to the students justify their fancier title: “Student Technology Consultants” instead of simply “lab monitors”. Colvin (2007) also lists the benefits of tutoring programs: elimination of hierarchical structures; increased motivation for students and tutors; and empowerment for tutors. According to Topping (1996), peer tutoring programs may result not only in intellectual or formal academic achievements, but also such projects may have “affective and attitudinal gains, social and emotional gains, self-image and self-concept gains, or any combination” (p.52).

At Miotti, the program was called “Student Monitor” and in its first edition it had as participants: 18 students from sixth to eighth grades, two teachers from the school and four researchers (from the fields of computer science, psychology and pedagogy). Initially, 68 students had signed up to join the program. The selection criteria were not based on the achievements of the candidates, but on their presence in the first meeting. This selection criterion, which was not based on students’ former knowledge or grades, might be a target of criticism from some educators. However, as Topping (1996) points out, even if there might be a traditional premise that tutors should be the ‘best students’, if tutors are average (or even less) students, both tutor and tutee may find cognitive challenges in their joint activities: “Although tutee gain may not be so great, the aggregate gain of both combined may be greater” (p.51). Nonetheless, most of the times the Student Monitors at Miotti were assigned to help younger children. "Supported (or 'scaffolded') exploration through social and cognitive interaction with a more experienced peer in relation to a task of a level of difficulty within the tutee's 'zone of proximal development' remains a theoretical cornerstone of peer assisted learning (Vygotsky)" (Tooping, 1996, p.52).

The Student Monitors had meetings with the researchers every two weeks. In these encounters, the students had the opportunity to share their feelings and understandings about their role. When asked about the most challenging experience in this program, one of the students reported that it was to work with handicap students. She was aware of the fact that the experience helped her to be more patient and to understand other people’s needs. Such kind of experience involve technical (the use of laptop), formal (observation of rules of behavior in social interaction, formal meeting, etc.), and informal (gathering with other monitors) aspects. The emotional and affective responses and outcomes transcend these categories, impacting these students life with valuable learning.
Discussion

The successful cases reported previously were not accomplished without difficulties. Even though the XO laptop had been specifically planned to be used in the context of developing countries, those who idealized the OLPC project probably were not aware of every situated challenges that may emerge in real life settings. Some of the challenges faced are commented here and summarized in Table 1.

Some teachers were reluctant to merge their practices with the new artifact, as it seemed a toilsome task for them to learn to operate the laptop. Even among those who were more familiar with new technologies, there were some more conservative teachers who preferred not to join interdisciplinary activities as they were not part of the official pedagogical plan.

The possibility of taking the computer home provides plenty of new opportunities for students and their families. For example, parents, siblings, neighbors may interact with the laptop, which contributes to the digital inclusion of the community. However, what we experienced in this school was the concern that some teachers and parents had about children’s safety. They feared that their children could be robbed in their way home. The concern of having the laptops stolen by own family members was also mentioned.

As not all children were able to take the laptops home because of the above mentioned problems, other problems arose: how to store those laptops at school? How to transport those laptops back and forth from the storage room to the classrooms? How to recharge the batteries of more than 30 laptops rooms with at most two wall sockets and not enough extension cords? As trivial as the solution might seem, the purchase of extension cords might not be so simple for most of the Brazilian public schools.

Another issue related to the technical layer was internet access. Miotti had broadband internet connection as well as wireless routers, which allowed students to do their jobs. However, the number of routers did not seem to be enough for students’ simultaneous connection: the internet connection was often slow and intermittent. The interruption in the flow of actions or thoughts caused by slow or intermittent internet connections might contribute to lowering the students’ levels of motivation and interest. Moreover, students often credit such failures to the laptop, instead of to the school’s internet network.

<table>
<thead>
<tr>
<th>Informal</th>
<th>Formal</th>
<th>Technical</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Most parents would not be able to use the laptop and participate in children’s activities;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Some teachers are not comfortable working with technologies they do not master.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Not all children were allowed to take the laptops home due to teachers and parents’ fear of theft;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Some teachers worried that interdisciplinary activities would not fit the official pedagogical agenda.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Difficulties due to the laptop’s response time;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Internet connection not always available;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Operational system has different interaction models;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Not enough wall sockets or power extension cords.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The way a new technology is inserted in the school and the way the issues that may emerge are handled might influence how users affectively respond to that technology. The approach adopted by Miotti, which allows the school’s culture and practices to influence the use of the new technology, is expected to contribute to more positive responses from students, as indicated by the preliminary results of our investigation (Hayashi et al., 2012a). Activities that promote positive affective responses are important in educational contexts. Some renowned authors like Piaget and Vygotsky had already pointed out the importance of affect in learning processes. For Piaget, affectivity is an energetic source: the fuel that makes the motor of intelligence run, without changing its structure (Devrie, 2006). For Vygotsky, affect and intellect should be considered simultaneously in the process of making sense of things (Oliveira, 1992).

The cases from the previous section were taken from the real and complex scenario of a school’s ecosystem. For each of those cases, positive emotional and affective outcomes were identified. Table 2 summarizes some of the outcomes. They represent just a few of many opportunities that can emerge from such scenarios.
Table 2. Emotional and affective outcomes identified in each case/learning opportunity

<table>
<thead>
<tr>
<th>Learning opportunities</th>
<th>Emotional and affective outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework at school</td>
<td>Collaborative learn-to-learn activity → teachers aware of learners’ emotional responses towards homework assignments</td>
</tr>
<tr>
<td>Interdisciplinary activities</td>
<td>Knowledge/information outside the boundaries defined by the label of a discipline → more motivated learners</td>
</tr>
<tr>
<td>Outside the school</td>
<td>Pride as indicative of the delight arising from the possession of the laptops → more motivated learners; impacts on self-esteem</td>
</tr>
<tr>
<td>Student Monitors</td>
<td>Learning about responsibility and social skills → impacts on self-esteem; empathy</td>
</tr>
</tbody>
</table>

Affective and emotional aspects were present in all interactions. Being aware of how affective responses take place as well as of the elements that triggers positive responses might help teachers and students to repeat those positive responses/experiences. That might contribute to more satisfying, productive and integrated learning opportunities. Therefore, we argue that not only the usability of a technological artifact should be considered in the design processes, but also its Affectibility (Hayashi & Baranauskas, 2011).

Conclusion

One of the challenges educational technology faces nowadays is to enrich the formal learning settings while maintaining a closer connection to the learners’ informal contemporary world. In this paper we brought to discussion the understanding of Information Systems’ formal, informal and technical layers of a framework, which may contribute to more holistic view of digital technology use in education.

We presented four different scenarios of XO laptops usage that illustrate how formal and informal learning might happen simultaneously. The cases shed light to the initial question “how can a new technology be implemented in schools in a way that it makes sense to the users?” Moreover, the combination of different practices allowed learning in different places (at home and at school) with different learning intentions (be them explicit or not) to be merged into more interrelated learning scenarios. Therefore, the cases suggest a positive response to the question “Could that technology contribute to more integrated learning scenarios?” The cases were a result of Workshops conducted in the scope of the XO Project (2010), and of Participant Observation on teachers and students daily practices.

Emotional and affective aspects were visible in all case scenarios, suggesting their need to be explicitly accounted during the design of technology for educational purposes.

References


Eraut, M. Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, 70, 113-136.


OLPC. http://one.laptop.org/ (last access: Jan., 2011)


Lecture Capture in Engineering Classes: Bridging Gaps and Enhancing Learning

Hasan Al Nashash and Cindy Gunn
American University of Sharjah PO Box 2666, Sharjah, UAE // hnashash@aus.edu // cgunn@aus.edu

ABSTRACT
This paper explores the use of lecture capture in Engineering classes to provide students with the opportunity to enhance their understanding of the course content. Students were asked to provide feedback on what they perceive the benefits and the drawbacks of lecture capture to be. The results show that the students consider lecture capture an effective tool to help them succeed in the course. The videos are available to them 24 hours a day, seven days a week thus allowing students to bridge the gap between what they have understood in the formal class setting and what they are able to better understand after reviewing the videos in a more informal, relaxed environment. In addition, most of the students indicated that the availability of the videos did not encourage them to skip or miss any classes. The main drawback was associated with technical difficulties which resulted in some wasted time.

Keywords
Lecture capture, Learner autonomy, Technology in higher education, Engineering

Introduction
Educause describes lecture capture as an “umbrella term describing any technology that allows instructors to record what happens in their classrooms and make it available digitally” (2008, n.p.) for the students to view after the class. According to Davis, Connolly and Linfield, “lecture capture has the potential to alter the way in which face-to-face teaching is delivered and received” (2011, p. 11). The recording of classes can be a simple video recording, but more often special software and / or hardware is needed so that the professor, the whiteboard and any other materials used by the professor, for example, PowerPoint, are recorded clearly. Well-known universities such as MIT and UC Berkeley were pioneers in lecture capture and webcasting (http://ocw.mit.edu/courses/ . http://webcast.berkeley.edu/).

Several recent studies have looked at the advantages and disadvantages of lecture capture from both the lecturers’ and the students’ points of view (Davis, Connolly and Linfield, 2010, Secker, Bond and Grussendorf, 2010, Toppin, 2011). Some of the benefits of including lecture capture are directly related to good teaching practice, i.e., fostering student engagement, appealing to students’ interests, offering multiple opportunities to access content and providing opportunities for learners to learn at their own pace. With lecture capture, students can view the archived lectures at their convenience and watch them as many times as they want or need to. The same can be said if the PowerPoint slides or other class notes are made available to the students, but lecture capture allows for the student to once again hear the Professor’s explanation and any class discussion of the content that occurred in class. The ability to both hear and see the information has been shown to help students retain more information and ultimately improve their class performance (Traphagan, Kucsera and Kishi, 2010). In addition, students can freeze the frame, rewind and hear the explanation many times until they feel they have understood the material.

One of the concerns expressed by faculty regarding the use of lecture capture is the possible impact on class attendance. Traphagan, Kucsera and Kishi (2010) looked specifically at this issue. They did find that the “availability of webcasts negatively impacted student attendance” (p. 19), however, the access to the video lectures and resulting better performance by the students “appeared to nullify the negative effects absenteeism” (p. 19) could have had. Toppin found in his study that there was no effect on student absenteeism and “actually increased attendance for a few students” (p. 392-393). In addition, in their review of recent lecture capture literature Secker, Bond and Grussendorf noted that “recording lectures has a minimal impact on students’ attendance” (n.p.).

Other concerns on the part of faculty include the extra work involved as well as privacy and copyright issues. Editing and publishing the videos can be time consuming. As noted by Educause, “a complicating element of lecture capture is ambiguity over who … owns the intellectual property once the recording has been made” (2008).

Students, for the most part, report that they benefit from having access to the video lectures. The benefits include higher grades, more active involvement in the class and engagement with the course material. In Toppin’s study the
“students valued VLC [Video Lecture Capture] as a supplement to their traditional lecture format and affirmed that it helped them to understand concepts taught in the courses” (p. 391). In addition, as noted by Zhu and Bergom, “students report that they appreciate the flexibility of accessing [videos] anywhere and anytime” (2010, p. 3). Students also state that having the video lectures available to them helped them focus more on the class because “they took fewer notes during class and were able to pay closer attention to the lecture” (Zhu and Bergom, 2010, p. 2).

Brooks, Epp, Logan and Greer (2011) looked specifically at how students were using lecture capture. They found evidence for four types of learners:
1. Minimal activity learners: students try watching the lectures but do not use it on a regular basis.
2. High activity learners: students may not watch all of the lectures but view some of the content at least once a week.
3. Deferred learners: students do not watch the videos until the end of the semester.
4. Just-in-time learners: students view the lectures only around midterm and exam time.

D’Angelo and Woosley (2007) observed that at the time of their study “students’ perceptions of technology use have not received widespread attention” (p. 464). Four years later, Brooks, Epp, Logan and Greer (2011) note that “very few studies have been done on the ways students use lecture capture technology to assist in their learning” (n.p.). This study, undertaken at the American University of Sharjah in UAE, looks at both of these issues. The study sought to answer the following three research questions:
- What benefits do students feel they gain from the webcast lectures?
- What drawbacks do the students see with including webcast lectures?
- How often are students viewing the lectures?

In the rest of this paper, we first offer a background to the study, describe the different technologies used, benefits of each technology, and some of the problems encountered. We then describe the data collection vehicles used to answer the research questions. After the data analysis, we discuss the findings and conclude with some ideas for further research.

**Background to the study**

This research was conducted at the American University of Sharjah (AUS) in The United Arab Emirates. AUS is a coeducational private university that was established in 1997. It is a small, Liberal Arts college offering both undergraduate and graduate programs with an approximate student body of 5,500 students from 84 different countries. The College of Engineering is the largest college, constituting almost 40% of the student population in both Graduate and Undergraduate classes.

The initiative of using video capture in Electronics course lectures for electrical and computer engineering students at the American University of Sharjah (AUS) started back in 2009. Since then, it has evolved through two main phases. In the first phase, the whole lecture was captured using a standard video recording facility. In the second phase, an interactive whiteboard technology was used which transformed the standard whiteboard into an interactive whiteboard.

In phase I, the lecture was held in the Engineering auditorium. The auditorium was equipped with a permanent desktop, an overhead projector of type Hitachi CP-X605 and a video reordering system of type Vaddio. The instructor used the computer for presenting PowerPoint lecture slides and relevant videos segments. The Vaddio video camera had a robotic tracking system where the camera was able to dynamically track the instructor’s movements while recording without the need to wear any tracking devices. The sound signal was picked up with a wireless clip microphone of type Shure. The video and sound data was instantaneously transferred to the audio-video recording system located in an adjacent room. This recording system was of type Hitachi. For every lecture, the following tasks were implemented:
- The IT technician ensured that the recording system was switched on and functioning properly
- At the end of each lecture, the IT technician burned a copy of the recorded lecture on DVDs and passed them over to the instructor
• The lecture material was then edited and compressed using available software such as MS Movie Maker. Each lecture required up to two hours of editing. After compression, a 75 minute lecture would be compressed down to 400MB of data.
• Due to the huge size of the recorded data, the instructor had to pass the edited lecture data to the library to save it on a special server for video streaming purposes.
• The library then informed the instructor of the data link to be used in the course website.
• The instructor added the link of the lecture video streaming to the course website on Ilearn and informed the students of its availability.

The above procedure was successful in capturing and webcasting all course lectures. Students had the opportunity to view the lecture videos 24 hours seven days a week. Nevertheless, there were some associated practical problems or obstacles faced by the instructor. These problems included:
• Learning curve which resulted in poor sound recording of the first few lectures.
• Poor lighting quality of the lecture hall as depicted in Figure 1.
• Instability of camera while tracking the instructor.
• No zoom function to enhance visibility of what was written on the board.
• The length of time needed for editing videos.
• Long delay from class lecture to videos being available for viewing by students.
• Lack of venue flexibility since the technology needed was only available in the auditorium.

Figure 1: Screen shot illustrating poor lighting

The problems detailed above prompted us to search for an alternative technology with better light quality, venue flexibility and less editing time.

Lecture capture technology

There were several systems available in the market to choose from which can meet our requirements. The system which we selected was the eBeam Edge system developed by Luidia Company. It is a Windows-based system with easy to install, purpose specific software aimed at creating an interactive whiteboard environment. This system is an interactive whiteboard technology which transformed the standard whiteboard into an interactive whiteboard. The system setup is detailed in Figure 2. It consists of the standard whiteboard, a data projector, a desktop, the eBeam edge transceiver and a stylus pen. The mobile transceiver which transfers the pen movement to the computer is a very light weight device the size of a board eraser and can be magnetically placed on the side of the white board. The
supporting software offers multiple formats, including jpg, pdf, and PowerPoint. It also enabled us to select different colors, different line thickness, and readily available shapes. The sound was also recorded using wireless microphone and was transferred directly into the computer. The sound system was also activated simultaneously when recording is activated which adds more flexibility to the system. The process of system setup and calibration would take around five minutes. Once the lecture is complete, the data was already saved on the hard disk of the computer. There was no need for any further editing or data compression. A 75 minute lecture would require around 30MB. There was no need of an IT technician to transfer data or send the data to the library for storage on their server. The instructor can easily upload the lecture file to the course website on Ilearn.

Figure 2: E-beam technology (http://www.luidia.com/)

There are many advantages of this system over the system we used in phase 1. We managed to reduce the lecture availability time from two days down to two hours. The quality of the video was superior as shown in Figure 3 because the room ambient light level was irrelevant in this case. The amount of data storage was also reduced by a factor of ten. The main disadvantage of this system was the inability of this system to video record the instructor while lecturing. This is a disadvantage because it deprives the student from learning from the body language of the instructor.
Results and Data Analysis

There were 40 students from two Electrical Engineering classes involved in this study. The course, ELE 241, Electronics I, is a required course for all Electrical and Computer Engineering students. 27 male students and 13 female students were involved in this study. Every lecture was captured using the e-beam system described previously and published for students on the AUS course management system, Blackboard. Blackboard is a course management system which allows adding all course elements online including course material, grades, email, announcements, assignments, etc. It also enables the instructor to generate course reports to display all user activity for all areas of the course, as well as activity dates, times and days of the week (http://ondemand.blackboard.com/).

Both qualitative and quantitative data were collected to answer the research questions via three venues: a student survey, focus group interview and the statistics from Blackboard. 38 students filled out the survey and four students, two men and two women, participated in the focus group. The Blackboard statistics represent all 40 students.

To answer the first question, *What benefits do students feel they gain from the webcast lectures?*, we asked the students in the survey specifically if they thought the videos were helpful in understanding the course material and if they thought the videos would help raise their grades. As show in Figures 4 and 5 below, 37 out of the 38 students either strongly agreed or agreed that the videos were helpful in understanding the course material and 34 out of the 38 thought having access to the videos would raise their course grade.

![Figure 4: Answer to question one in the survey](image)

![Figure 5: Answers to question 2 in the survey](image)
In the open-ended question of the survey, students elaborated upon the benefits of these two main points as illustrated in the following three quotes, none of which have been edited, from students:

The video lectures allowed me to study as efficiently as possible. I am taking the information straight from the professor. It also eliminated the need to ask the professor about small points because you can always go back to the video.

Since it is a class of 1 hour 15 minutes, sometimes it is quite difficult to give our attention throughout the class. If there is something we do not understand in class or miss out, these video lectures come in handy.

In addition to the usefulness of the videos to “fill the gap” if a student gets distracted in class for one reason or another, in the focus group interview, one student mentioned the ability to review to the lectures anytime and anywhere.

Although as reported in the literature, many faculty are concerned that the use of video lectures will encourage students to skip classes, the majority of the students who participated in this study indicated the opposite was true. As shown in Figure 6, when asked if the video lectures prompted them to skip class, seven students were neutral, three students said the videos do prompt them to skip while 28 either agreed or strongly agreed that the video lectures do not prompt them to skip classes.

In the open ended question one student noted,

This lecture video is very helpful because it will help you understand and see and visualize the lecture. It will not prompt me to skip, because you watch it again to get the things you missed during the class due to daydreaming or so. In addition, it helps you to organize.

Two students noted that they found the videos to be helpful when they missed classes. As they point out, students sometimes have to miss a class for various reasons. In cases such as this, the student has access to the videos to keep him/her from falling behind in the class:

This technology was very helpful when there was reasons I needed to skip the class for (strong reasons) but I did not watch the videos otherwise. And it did not prompt me to skip classes because I care about attendance and attending the class is still better.

The video lectures have definitely helped me a lot. I had missed out on a few critical lectures and was having a really hard time in understanding the concepts and analytical tools in the text. But because of the lectures posted online, I was able to understand; additionally, when I went back to the book
AFTER listening to the lectures, I was able to grasp that information too. So the lectures helped me greatly.

When asked if the students thought other professors should use this technology in their classes, 100% strongly agreed that they should. They gave the following reasons:

Some people have problems with remembering what the professor explained in lecture or don't know exactly how to take notes for it. The video helped me a lot in revising and studying for quizzes and exams. It will be great if all professors use this technology for all courses.

The lecture videos are very useful and it's an advantage to students. Because students can access the lecture video if they have missed some points in class or to understand things if something wasn't clear. All in all, it is a great technology, and it would be better if all instructors used this technology.

I would strongly recommend this technology to be used by other professors in other courses. It is very helpful.

To sum up the answer to our first research question, the data from the survey and the focus group indicated the students could see many benefits but the main reason for viewing them was linked to gaining a better understanding of the material, which, in turn, gave them the possibility of increasing their grades. Another benefit regarding the availability of the video lectures is that for some students, they are freed from taking copious notes in class and can concentrate more in the formal setting of the lecture. Most of the students in this study see having access to the video lectures as another tool to help them learn the course content, not as a replacement for, or an excuse to skip, the lecture.

To answer the second research question, What drawbacks do the students see with including webcast lectures?, we asked the students two questions on the survey, specifically, Were the lecture videos easily accessible? and Did you have any problems while loading the videos? These questions stemmed from previous requests for feedback from students in 2009 and 2010 regarding the use of video lecture capture in their classes. As shown in Figures 7 and 8 below, the majority of the students did not have any problems accessing or loading the video lectures. However, in the open ended question of the survey, students noted that the drawbacks of this system have to do with the time taken from the lecture when there is a technology “glitch”. As one student commented, “Sometimes, it gets stuck and you need to reset it in order to work and takes time from the lecture.” This point was also noted in the focus group interviews as well as the problem of the microphone not being turned on or running out of batteries. Another drawback of including the videos was presented by a male student in the focus group. As he noted, “It’s like seeing the lecture twice. if you didn’t like it the first time you won’t like it the second time!” This aspect of liking the material (and possibly the Professor) and its impact on learning is outside the scope of this research but is an area for further research.

![Figure 7: Answer to question 5 on the survey](image-url)
As noted above, the majority of the students said that having the opportunity to view the lectures did not impact their decision to attend class or not and, for the most part, the students find the video lectures helpful in their understanding of the class material. However, we wanted to match their perceptions and comments to their actual usage. Thus, to answer the third research question, *How often are students viewing the video lectures?*, we referred to the statistics reports from Blackboard platform. Figure 9 was extracted from one of these reports which shows a pie chart of the overall summary of students’ activity during the first 14 weeks of the semester. Students’ activities included course content, grades, email, etc. The figure shows that the largest activity area representing 72.9% was
occupied by visits to course contents which primarily contain the lecture videos. This is followed by grades which occupied 19%. This result indicates that students mainly visited the course website to watch the course videos.

The daily user activity of the course content is depicted in Figure 10. It shows rhythmic and regular visits by students to the course content. The peaks were observed in the days prior midterm exams. Otherwise, viewing the course contents reached a steady state after week seven of the semester and maintained regular visits to the course contents area. These statistics indeed support students’ claims that the video lectures were helpful in their understanding of the class material.

![Figure 10: Daily user activity of the course content on Blackboard](image)

The answer then, to the third research questions, How often are students viewing the lectures? is that the majority of the students are watching the videos every week. The students fall under Brooks, et.al., (2011) high activity learners category.

**Discussion**

The use of video lectures is an example of effective use of technology to promote both formal and informal learning as well as offering students the opportunity to take control of their learning. Research has shown that most people have limited attention spans. However, in a formal learning University environment, a normal class schedule can range from 50 minutes to three hours. If students lose their concentration or get distracted in class, the video lectures will allow them to view the lectures in a less formal setting, any time and any place that is convenient for them. One female student in the focus group said she likes to watch the videos during her commute home from University. Another male student mentioned that he has the video lecture running in the background while he is surfing the web or otherwise using the computer.

The availability of the video lectures also promotes learner autonomy. The use of Blackboard and viewing of the video lectures is something the students have full control over. Although their visits to Blackboard can be tracked, the students involved in this research were not given any incentives, such as bonus points, to do so. In addition, they can watch one part of the lecture as many times as they like and they can skip to any part of the lecture that they wish to review.

For most of the students involved in this study, having access to the video lectures had no impact on their face-to-face class attendance. They took advantage of the formal learning opportunities provided in the lecture and then supplemented their learning by viewing the lectures in their own time. However, for one student, the video lectures may have been the difference between passing and failing the course. When asked by his Professor why he was not attending class he explained that it was not just one class that he was not attending. He found it difficult to concentrate in all classes and stopped attending most of his lectures. He then told the Professor that he could keep up with his class because he had access to the video lectures.
Conclusion

D’Angelo and Woosley (2007) found in their study of students’ perceptions of educational technology “that students do not always see that technology benefits their learning” (p. 470). In this study, we have found, however, that most of the participating students could see the benefits of using video lectures. This study, unlike D’Angelo and Woosley’s study, focused on just one use of technology so perhaps the students were better able to focus on the advantages.

To sum up, this investigation shows that the students consider lecture capture an effective tool to help them succeed in the course. In addition, most of the students indicated that the availability of the videos did not encourage them to skip or miss any classes. The main drawback was associated with technical difficulties which resulted in some wasted time. This drawback can be rectified by introducing a preventative maintenance program by the IT department. Further to the data collected from the students’ surveys, the course management system’s reports indicate that students did indeed regularly view the course video contents.

From this research, we have identified two areas that require further investigation. At present, at AUS, there is only one professor using video capture with his classes. The students in this research all noted that they would like all of their professors to include video lectures. If this were to happen, an area of further study would be to once again investigate the students’ attitudes towards the inclusion of video capture in all or most of their classes to see if they value and benefit from the video lectures as much as they do now.

Using video lectures in class requires discipline on the part of professors. They must be willing to not only record each lecture but make sure that each lecture is available to the students in a timely manner. Professors, as well as students, can benefit from the use of video lectures, but as Secker, Bond and Grussendorf (2010), note there are very few studies done focusing on professors’ use of and attitudes towards video lecture. They further note that, “those who do exist tend to focus on what lecturers do not like about lecture capture” (2010, np). Thus, an area of further study would be an investigation into why or why not professors use lecture capture in their classes.

References


Animating Civic Education: Developing a Knowledge Navigation System using Blogging and Topic Map Technology

Tien-Chi Huang1 and Chia-Chen Chen2*

1Department of Information Management, National Taichung University of Science and Technology No.129, Sec. 3, Sanmin Rd., Taichung City 404, Taiwan // 2Department of Information Management, Tunghai University, No. 181, Sec. 3, Taichung Port Rd., Taichung City 40704, Taiwan // tchuang@nutc.edu.tw // emily@thu.edu.tw

*Corresponding author

ABSTRACT

With the advent of Web 2.0 technology, message transmission has become increasingly convenient, and the rising amount of information has become gradually diverse. A question must be asked of this trend, on whether informal learning resources can be integrated into formal learning knowledge. This study attempts to integrate educational blog articles and use the wealth of information inherent in the development of a blogging system with knowledge navigation demand for both formal and informal learning in civic education. In the system, Topic Maps (TM) technology was implemented to represent informal learning content in a formal curriculum structure. To develop a framework adapted to formal and informal learning contexts, this study proposes a seamless model for teachers and system developers to represent informal learning content that adheres to the formal knowledge structure. Analytical results indicate that students approved the effectiveness of combining the blogging system and individual Topic Maps (iTM) function. In other words, the system derived from the design framework helps students learn citizenship courses in an innovative manner. The outcome further supports the proposed seamless model in providing an appropriate viewpoint for integrating formal and informal learning.

Keywords

Blog, Topic maps, Knowledge navigation, Civic education, Formal learning, Informal learning

Introduction

Civic education in formal education comprises four wide-ranging fields, namely, politics, law, economics, and psychology/sociology, fields of knowledge that are crucial to becoming an ideal citizen. However, exam-oriented education in Taiwan limited civic education to textbook teaching in secondary education. Although textbooks used in formal curricula are valuable resources, solely focusing on written text without establishing a connection to real-life situations decreases the relevance of information in students’ lives. Therefore, this study proposes the online medium as an information source to eliminate the disadvantages of using formal learning content in teaching and learning civic education (Du & Wagner, 2007).

Although web-based learning has been widely used for several years, teachers were regarded as the primary source of learning materials (Huang et. al., 2009). Since the advent of Web 2.0 solutions in 2004, numerous user-centric information platforms and applications in the form of blogs, Wikipedia, and YouTube have expanded widely. Among these resources, blogs are the most appropriate educational medium because of their low threshold, ease of use, and incorporation of various resources including multimedia, individualization, and socialization characteristics (Kim, 2008). Blogs have proven to be an adequate teaching platform for students in higher education (Kang, Bonk, & Kim, 2011; Huang, Huang, & Yu, 2011). Their informal interactive nature encourages students to direct attention to specific aspects, reflect on certain concepts and ideas, and eventually receive feedback (Bransford, Brown, & Cocking, 1999). The main advantage in using educational blogs is that they can be an ideal interface not only for delivering formal learning content, but also for reporting informal life experiences, news, and information worldwide (Huang, 2011; Lin & Kao, 2010). However, because of the information and knowledge scattered over numerous blogs, learners can experience difficulty integrating and absorbing self-collected learning materials without a systematic approach. Thus, a certain level of systematization is required to ensure learning effectiveness. In other words, educational blogs should integrate navigation functions to represent information in a systematic manner, to optimize efficiency. Therefore, this study develops an educational blogging platform equipped with a knowledge navigation function to animate civic education between students and teachers inside and outside the classroom.

We used an ISO standard named “Topic Maps” to structure the knowledge representation of the “Citizenship and Society” course and to link the corresponding blog entries as supplementary learning resources. The design of the
knowledge structure follows the Newly Revised High School Curriculum Guidelines (99 Guidelines) issued by the Ministry of Education in Taiwan, and schedules the crucial knowledge sphere, concepts, and issues in every subject domain in the form of a chapter or textbook captions.

**Background**

**Civic education and Informal learning**

A free society must ultimately depend on its citizens, and the approach for infusing people with necessary qualities is through education (Branson, 1998). Civic education is a primary and important component in developing holistic education. The goal of civic education is to cultivate positive values and attitudes in students, to discharge their responsibilities with the necessary knowledge and skills. To cultivate a citizen, an objective of civic education is to link learned knowledge with daily life experiences. However, civic education in secondary education in Taiwan is often limited to classroom teaching because students are often unable to experience the usefulness of acquired knowledge in their daily lives. This subject should deserve more than a passing notice. Whereas considerable attention has focused previously on research issues related to the development of civic education curricula in schools, literature on the development of informal civic education has emerged slowly and has been more scattered. In light of these concerns, an approach to bridging classroom teaching and daily life experience is necessary.

Past studies have identified five primary characteristics of informal learning. (1) outside educational establishments (Sung, et al., 2008, 2010); (2) originates accidentally and sporadically, in association with certain occasions, from changing practical requirements (Schugurensky, 2000; Sung, et al., 2010); (3) related to situation management and fitness for life (Clough et al., 2008; Janssen, Berlanga, & Koper, 2011); (4) experienced directly in its "natural" function of everyday life (Gu, Gu, & Laffey, 2011); and (5) spontaneous (Janssen, Berlanga, & Koper, 2011). These characteristics form the basis for this study to identify a research niche for developing informal civic education, and the inspiration to adopt an innovative technology to achieve it. Blogs, a Web-based technology, have been often adopted to disseminate useful information in educational contexts, particularly in informal learning contexts (Park, Heo, & Lee, 2011). Hence, this study uses blogging technology in an attempt to combine class teaching with informal civic education.

**Blog roles in educational contexts**

A weblog (blog) is a Web 2.0 technology that allows people to share their thoughts and comments quickly with the entire online population. People unfamiliar with Web design codes (HTML, CSS) are successfully able to post an article with multimedia materials. This advantage has recently gained the attention of e-learning literature to pedagogical roles. The literature is filled with discussions surrounding the pedagogical usage of blogs, which can be categorized roughly into two types of roles: formal learning usage and informal learning usage (Guazzaroni et al., 2010). Blogging opens possibilities for representing formal or informal learning content in an informal learning context, and blogs often bridge formal and informal learning. Formal learning content includes structured and organized content ordered by date or curricula categories, whereas informal learning content typically involves learning experiences in daily life. Blogging creates new possibilities for self-expression and social interaction with peers. Writing blogs not only relieves emotional stress and stimulates self-expression and reflective thinking skills (Xie, Ke & Sharma, 2008), but also enables networking and sharing of resources and ideas in a professional learning community (Farmer, 2004). Reading blogs also facilitates self-reflection (Brescia & Miller, 2006) and enhances community membership and participation (Baumer, Sueyoshi, & Tomlinson, 2011). Blogging plays a synthesized role in providing numerous opportunities for self-reflection and peer interaction.

A blog is also suitable for managing an online learning portfolio because of its permalink and reverse chronological order features. The permalink is a URL that indicates a specific blog entry to enable bloggers to maintain their learning processes as a longer-lasting URL for reference (Treese, 2004). These permalinks can be archived by date or categories named by bloggers to provide an enhanced approach to represent the learning portfolio. Users can easily review his or her learning experience represented by blog entries displayed in reverse chronological order. The learning experience typically includes generating personal learning goals, planning how to tackle a problem,
implementing the means for tackling a problem, evaluating whether learning goals have been met, and replanning based on this evaluation.

Although blogging opens possibilities for representing formal or informal learning content in an informal learning context in literature, attention has been minimal in how to fully integrate informal learning content with the formal curriculum structure on blogs. Therefore, this study develops a blogging system for supporting both formal and informal learning in an informal learning setting. Accordingly, new approaches should be devised on how to acquire knowledge and skill, and how to efficiently deploy learning resources that are constantly updated with the current knowledge economy (Guazzaroni et al., 2010). However, the rapid increase of blog resources and modality types has led to a great accumulation of information (Huang, Huang, & Cheng, 2008, 2009). Another important issue is the development of an approach for extracting knowledge from the massive unorganized information resources online.

**Topic Maps and knowledge navigation**

This study used the information navigation system to address these issues by constructing information. Although this type of navigation system solves the issues, most existing systems have shown that information matching and keywords lead to incomplete outcomes and lack expansibility in knowledge representation.

Therefore, this study proposes the latest concept in this domain, “knowledge navigation,” to overcome the incomplete issue. Based on semantic Web technology, knowledge navigation connects customers and resources more effectively by presenting information resources for users within a specific knowledge structure (scope) and providing a search service. This method accounts for both the resource view and the knowledge view in providing users with more diverse, qualitative, and complete information.

This study used Topic Maps (TM), known as the “GPS of the information universe,” which is an ISO standard, to build a blogging system with knowledge navigation. Formulated in XML-syntax, TM possesses interchangeable features that allow it to describe the relativeness between concepts and even link it to online information resources easily (Pepper, 2010). Three primary elements defined in TM, topic, association, and occurrence are also adapted to knowledge system development (Rath, 2003). Therefore, in online education settings, TM is an appropriate method for instructors to demonstrate subject knowledge and connect it to massive online resources in an organized manner. Assisted by TM, users and students can experience information navigation more efficiently. XML Topic Maps (XTM) have become a mainstream of knowledge navigation and are applied in many fields (Stefan & Ludwig, 2002; Petra & Helmut, 2005).

Zhai, Wand, and Lv (2008) applied XTM to develop an urban traffic information portal for knowledge navigation. By using topic map query language, users can retrieve related information on urban services, such as hotels, banks, and hospitals. Wu and Wang proposed a method to build a TM-based knowledge navigation model for explicit knowledge navigation. The results showed that TMs offer an appropriate approach to visualize explicit knowledge and to enhance our understanding of knowledge navigation for the customer services domain of mobile communication corporations in China. Likewise, Kemény, Erdős, and Váncza (2008) successfully used topic maps to enable usability and accessibility of knowledge resources of a research community to members of the group.

This literature review shows the application of knowledge navigation in various domains. However, empirical studies on TMs for knowledge navigation in schools remain relatively limited. Current educational blogs are not typically customized for educational purposes for user interface and function features (Kim, 2008). Previous empirical studies have not emphasized the usage and sharing of knowledge values in blog entries. To develop a blog-based knowledge navigation system (BKNS) for creating a new role of blog usage in schooling, this study uses TMs as a tool for representing knowledge guided by the proposed framework and conducts an empirical study to determine the extent to which the blogging system can construct and navigate knowledge.

**The proposed framework**

The BKNS provides a personalized service to individual students based on their learning portfolios. This section details the system architecture, user roles, components, and TM designs.
System architecture and user roles

Figure 1 shows the BKNS architecture, and the numerals indicate the system operation procedure. User roles include the teacher and the students. In the BKNS, the teacher is responsible for designing the TM framework, and the students are responsible for posting and reading blog entries as occurrences in TMs. The system comprises three repositories: the blog knowledge base, the topic association repository, and the TM repository. Several operations are conducted during the learning progress. First, the teacher creates the TM framework (Step 1). During the topic association creation process, the teacher might refer to the types of course materials, such as textbooks, multimedia resources, online resources, and academic papers. The topic association schema is the backbone of TMs, which defines topics in a certain course and associations between topics, stored in the topic association repository. Second, a blog knowledge base stores all posts published by the teacher and students. Finally, the TM repository maintains all topic map files (XTM format).

Students have no limits to access the blog entries posted in the BKNS, but only registered students can publish blog entries in their respective blogs and obtain an individual knowledge navigation service. After students log into the BKNS, the blog-authoring module offers a What You See Is What You Get (WYSIWYG) editing environment for publishing blog entries (Step 2). This module allows students to view information similar to the result while creating the posting. After the registered learner publishes his/her own post, or reads posts published by others, an e-portfolio detection agent records a blog entry status (Step 3). Blog entries read or written by students are considered occurrence links in the students’ own TMs, and a topic mapping agent embeds these links (occurrence syntax) into the existing TM (Step 4). Therefore, all students have their own TMs to represent the individual knowledge structure. These TMs are generated by the TM-generating module and stored in the TM repository (Step 5). Finally, a TM is parsed by the TM parsing and representing module, and is represented graphically in radial form for individual knowledge navigation (Steps 6 and 7).

![Figure 1. The system architecture of BNKS](image)
System components

Blog authoring module

This module provides an interface for users to post a blog entry. The basic elements in a post such as title, category, date, tags, and post body are editable by using this module. This module uses a WYSIWYG editor for easy embedding of images, videos, text decorations, and hyperlinks in a post, which infers that the editor has the ability to directly manipulate post layout without having to type or remember names of layout commands (HTML codes). This function also enables users to emphasize certain aspects of a blog entry by using text decorations such as underline, bold, italics, and colors.

![Figure 2. The blog authoring module](image)

E-portfolio detection agent

This agent records each post that the teacher and students write and read. When a user reads a post over 20 seconds, the agent determines whether the user has studied the post and marks the post as “a read post” in a specific category. When the learner posts a blog entry, the agent marks this entry as “a written post” and records its author information into the system.

Topic mapping agent

The topic mapping agent is responsible for linking posts to corresponding topics. This agent embeds the XTM codes into the TM file when a user posts a blog entry in the blog knowledge base.

Topic map-generating module

The TM-generating module generates XTM files for each user. Because every BKNS user can have his/her own written/read posts, the TMs (XTM files) generated by this module are individualized and unique. After the TM designer (the instructor in this study) builds the schema of a TM, this module organizes all XTM codes and generates
the first XTM file. When a user registers an account and posts blog entries, this module automatically generates an independent XTM file for each user.

**Topic map parsing and representation module**

The topic map parsing and representation module is responsible for parsing each XTM file and representing it graphically. The following figures show two examples of graphical TMs, which are centered by the political concepts of “Polity and Form of Government” and “Democratic Theory.” Each node in the figure can add extra information such as colors, an introduction, and other links related to this topic. The functionality of these graphics is similar to that of the Belvedere system, designed to teach science-related subjects to high school students unfamiliar with the science domain and to recognize abstract relationships in scientific theories (Suthers et al., 1995). Instead of using reason with students, this module imparts in-depth knowledge of what they have read and written. According to the theory of human mental ability, people can process and remember visual information quickly (Gordin & Pea, 1995). Therefore, visual representations of information and knowledge should aid people in learning.

![Figure 3a. Topic map example centered with "Polity and Form of government" topic](image)

**Figure 3a. Topic map example centered with "Polity and Form of government" topic**

![Figure 3b. Topic map example centered with "Democratic theory" topic](image)

**Figure 3b. Topic map example centered with "Democratic theory" topic**

**Topic map design**

Before users can use the BKNS, the TM designer must build the TM schema, which is the backbone of the TM and is designed according to the subject concepts. The standard process of achieving TM design and implementation is outlined in the following steps:

- **Step 1: Define the application domain**
  This study defines "Citizenship and Society," a social science curriculum in senior high school, as the application domain. This course in senior high school encompasses the four main subjects of psychology/society, politics, law, and economics. Because it is a social science field instead of a literary field, it is characterized by methodology and a knowledge structure, and as a social science, real events and real people comprise the subject's features. In a TM, topics and associations construct a knowledge structure by presenting concepts and their relations, whereas occurrences indicate real events, objects, and items (linkable blog entries) to help students transfer learning.

- **Step 2: Define the functional requirements**
  In this study, the TM plays the role of an instructional/learning tool integrated into the developed blog system. Therefore, the main users are citizenship teachers and students. All students have their respective TM space to
present topics and associations set by the instructor and occurrences filled with blog entries posted by students. Therefore, every student has his or her own TM through the designed individual topic map function.

- **Step3: Define the schema**
  Issued by the Ministry of Education in Taiwan, the Newly Revised High School Curriculum Guidelines (99 Guidelines) delineated the crucial knowledge sphere, concepts, and issues in every subject domain in the form of a chapter or textbook captions. This study presents these concepts and issues of "Citizenship and Society" by using the TM. For instance, "Prime Minister" and "Monarch" are two major concepts in the "cabinet system," and these three concepts (Prime Minister, Monarch, and cabinet system) are topics in a TM. Associations present the relations between topics. For this case, the association between "Monarch" and "Prime Minister" is "to nominate."

- **Step4: Select tools and implement application**
  The blog is used to implement the application. As mentioned in the literature, blogs are an efficient teaching/learning tool for combing multiple online resources. To enforce its function in presenting structural information, this study developed a blogging system to aggregate student learning portfolios, including text, images, and video content. These contents are considered occurrences in TMs.

- **Step5: Populate the topic map**
  As in the previous step, student posts are occurrences in TMs. Before a post is published, every student is asked to classify it into a specific category of the topic defined in a TM. Although misclassification may occur and cannot be checked by the system automatically, students can still use a review function to understand their errors for post reclassification. The curriculum expert who defines the TMs should also adjust the TM structure based on class instruction. A TM can be populated automatically by students and manually by the TM designer.

- **Step6: Maintain the topic map and its application**
  To keep the application running and the TM updated, students are asked to read and write blog entries in line with current events and news. As mentioned, the structure of a TM is maintained by the curriculum expert who defines the TM schema.

*Figure 4. The developed blogging system*
System demonstration

Figure 4 shows the developed blogging system applied to the subject of citizenship education. Students can post their learning content and interact with one another by posting comments. As shown in the figure, multimedia content (images, photos, videos) can be included in a post. When students edit a post, they can use the text-decoration function (bold, italics, colors, size, and fonts) to draw attention to parts requiring emphasis. In order to strengthen individual and adaptive learning (Wang, Wang, Huang, 2008), students can also construct their own knowledge structure using the “iTM” function representing **individual Topic Maps**, as shown in Figure 5.

![Figure 5. Individual knowledge representation using topic map technology](image)

The individual TM includes four sections: the main topic section, the topic structure section, the occurrence list section, and the visualized TM section. As mentioned, four main topics in the subject of “Citizenship and Society” (politics, economics, psychology, and law) are presented in the main topic section. The topic structure section uses a tree structure to present the topics in a selected main topic (politics is shown in Figure 4). Content in the topic structure section correspondingly changes when a different main topic is selected. The occurrences in each topic in the topic structure section are presented in the occurrence list section, and these occurrences are blog entries classified into two types of occurrences: read and written blog entries. As stated in the function of the e-portfolio detection agent, when a blog article is read by a student for over 20 seconds, it flags this article as a read entry for him/her. Similarly, the agent flags an entry as a written entry when a student posts an article on his/her own blog. Finally, a graphical TM is presented in the visualized TM section by TM parsing and the representation module. Students can easily view the TM presented in a topic-centric manner. By clicking other topics, they can change the TM display to understand the relations between two topics. In addition, every topic provides additional information (e.g., topic introductions and hyperlinks to outside resources), which can be provided in advance by the TM designer.

Methods

Experimental design

A three-stage experimental design was chosen to examine the effectiveness of the proposed framework. The participants were 48 second-grade students recruited from a senior high school in Taiwan. All participants received the same three-stage treatment explained below.

- 0/0 stage (teaching in the classroom without the aid of the BKNS)
- 0/1 stage (teaching with a blogging system)
- 1/1 stage (teaching with the BKNS).
Measures

Formal learning tests

To investigate learning efficiency facilitated by the blogging system and iTM, students were required to undergo three tests at various points during the semester. After completing a five-week course of Citizenship and Society (the 0/0 stage), students were administered the first test. After the first test, the students completed another five-week course (the 0/1 stage) before taking the second test. Finally, after the second test, the students completed another five-week course (the 1/1 stage) and took the third test. All the tests were paper-and-pencil tests with multiple-choice questions. Because the scope of the tests varied according to the course content, student scores (of four classes in the same grade) were standardized to determine whether the relative position among students varied at different stages. By using the following formula, student scores for the three tests were standardized to obtain measures of relative position of each student (i.e., the z score).

\[ z = \frac{X - \mu}{\sigma} \]

where X is a raw score to be standardized; \( \mu \) is the mean of the population (entire eleventh grade, 248 students); and \( \sigma \) is the standard deviation of the population.

Questionnaire: Effectiveness of the blogging system and iTM function

To examine student attitudes toward the knowledge navigation system, we developed a questionnaire containing 10 items. To measure student responses, we administered a questionnaire employing a 5-point Likert scale, ranging from 5 (strongly agree) to 1 (strongly disagree). The questionnaire focused on student attitudes toward the blogging system and iTM function. We verified that the questionnaire was valid and reliable because the two investigated factors explained 68.08% of the overall variance. For the validity, the coefficients \( \alpha \) for the internal consistency of the two factors was .85 and .83, respectively. After the time course of the blogging system and iTM, the questionnaire was administered to students after the second test and before the third test.

Results

Formal learning tests

The mean standardized scores and mean raw scores of the students for the three tests are shown in Table 1. For the mean raw scores, the results show that the mean raw score was lower for the second test, which was conducted after the students completed the course using the blogging system. By observing students’ standardized scores to determine their relative position, we found that the mean standardized score increased after the students completed the course with the iTM function.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Mean</th>
<th>Z score mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/0</td>
<td>81.2</td>
<td>69.24</td>
</tr>
<tr>
<td>0/1</td>
<td>79.8</td>
<td>69.79</td>
</tr>
<tr>
<td>1/1</td>
<td>84.1</td>
<td>72.94</td>
</tr>
</tbody>
</table>

Effectiveness of the blog and iTM

For student attitudes toward the blogging system, the mean rating of the students at the 1-1 stage was higher than that at the 0-1 stage, as shown in Table 2. By considering the results of the unstructured interviews, we found that students changed their attitudes toward the blogging system after experiencing the blogging system with the iTM function. Therefore, we infer that the students generally had a more positive attitude toward the iTM function than toward the blogging system alone.
Table 2. Students’ responses to the blogging system and iTM function

<table>
<thead>
<tr>
<th>Stage</th>
<th>Toward the blogging system</th>
<th>Toward iTM function</th>
<th>Toward the blogging system with iTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/1</td>
<td>Mean 3.24 S.D. 0.72</td>
<td>Mean 4.32 S.D. 0.61</td>
<td>Mean 4.56 S.D. 0.75</td>
</tr>
<tr>
<td>1/1</td>
<td>Mean 3.59 S.D. 1.14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*t = 3.06*

*p < .005

Correlation analysis

Correlation analyses of students’ raw scores for the three tests and their attitudes were performed. As shown in Table 3, students’ raw scores were positively correlated with their attitudes toward the blogging system (r = 0.51 for the second test; r = 0.52 for the third test) and significantly positively correlated with the iTM function (r = 0.75).

Table 3. Correlation analysis between achievements and effectiveness

<table>
<thead>
<tr>
<th>Prior achievements</th>
<th>Blogging system effectiveness</th>
<th>iTM function effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(current semester)</td>
<td>First test (0/0 stage)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second test (0/1 stage)</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Third test (1/1 stage)</td>
<td>0.52</td>
</tr>
</tbody>
</table>

For the correlation between students’ prior achievements and their attitudes, we found that students’ prior achievements were negatively correlated with their attitudes toward the blogging system (r = -0.73) and the iTM function (r = -0.68), indicating that the prior achievements of students was negatively correlated more positive attitudes toward the blogging system and the iTM function.

For the rating distribution, the standard deviation was close to 1, indicating a substantial variation among students. Therefore, this study categorized the students into groups to identify the cause of the variation. The students were first categorized by gender. However, no significant difference was found between male and female students.

Subsequently, the students were categorized into “prior-high” and “prior-low” groups based on their prior achievements. Students with scores for the previous semester in the top 50% were categorized into the “prior-high” group, and the remaining students were categorized into the “prior-low” group. By comparing the two groups, we found that the students in the “prior-low” group had a more positive attitude toward the effectiveness of the blogging system, compared to that of the students in the “prior-high” group. The same pattern was found for the two groups on student attitudes toward the iTM function.

Table 4. Group analysis at 1-1 stage

<table>
<thead>
<tr>
<th></th>
<th>Blogging system effectiveness</th>
<th>t-test</th>
<th>iTM effectiveness</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean S.D</td>
<td></td>
<td>Mean S.D</td>
<td></td>
</tr>
<tr>
<td>(a) Male</td>
<td>3.42 0.51</td>
<td></td>
<td>4.31 0.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female 3.63 0.49</td>
<td></td>
<td>4.11 0.65</td>
<td>p = .34</td>
</tr>
<tr>
<td>(b) Prior high</td>
<td>2.63 0.58</td>
<td>19.32*</td>
<td>3.46 0.51</td>
<td>14.59*</td>
</tr>
<tr>
<td>Prior low</td>
<td>4.56 0.57</td>
<td>*p &lt; .005</td>
<td>4.80 0.41</td>
<td>*p &lt; .005</td>
</tr>
</tbody>
</table>

Discussion

Creating a seamless context for formal and informal learning

The objective of this study was to investigate the efficiency of formal learning through an informal interface, such as a blog. Because the paper-and-pencil tests were developed based on a formal curriculum, this study evaluated the efficiency of formal education by using results from the three paper-and-pencil tests. By calculating the mean
standardized scores for analysis and comparison, we found that students’ standardized scores improved as the experiment progressed. For example, by using the blogging system and the iTM function, the students’ relative positions improved, compared to the earlier stage.

Further analysis of the correlation between student attitudes toward the blogging system and formal learning revealed that, when students had limited experience with the educational blogging system, an intermediate positive correlation emerged between their test scores and their attitudes toward the effectiveness of the blogging system. During the informal unstructured interviews, the students reported the degree to which they enjoyed informal learning using the blogging system. In addition, because the blogging system provided diverse course-related but extracurricular information related to daily life, the students could explore this information in relation to the courses or by sharing their life experiences.

However, this study has three limitations. First, it is the first to attempt using a blogging system to assist formal learning; thus, the ideal method for enhancing formal learning by using a blogging system has yet to be determined. Second, including extracurricular knowledge in addition to the knowledge presented in classes increased the knowledge that students were required to learn. Third, the learning environment was informal; thus, if the students logged into the blogging system from home, some may have been unable to relate information on the blog with the formal courses. Consequently, although the blogging system enabled the presentation of formal teaching materials in an informal environment, assistance from other tools is still required to fully integrate formal and informal learning.

To address the poor organization of the blogging system, which hinders its ability to enhance formal learning effectively, this study implemented the iTM function in the blogging system. By investigating the correlation between the iTM function and formal learning, we found that student attitudes toward the effectiveness of iTM were significantly and positively correlated to their test scores ($r = 0.75$). This pattern indicated that, with the assistance of a blogging system and iTM, students can effectively integrate formal and informal learning in an informal setting.

The influence of other factors on the seamless model

To understand whether other factors influence the seamless model, we further considered the gender and prior achievement categories. To investigate the influence of gender, students were divided into boys and girls. To investigate the influence of prior achievements, the students were divided into a prior-high and prior-low group, according to their prior achievements.

The analytical results showed that data variations can be attributed to whether a student’s prior achievements were high or low. In other words, lower prior achievements of a student indicated more positive attitudes toward the blogging system and the iTM function. Related literature suggested that, in addition to family background (Chang, 2009), students’ low motivation or interest to learn formal content they consider “boring” or “irrelevant to real life” is another significant factor in their low achievements, irrespective of whether they learned in class or over the Internet (Tsai, 1989; Bernard et al., 2004). Unstructured interviews with students in the prior-low group were conducted. The results revealed that students in the prior-low group had the following characteristics: preferring to learn through various teaching methods, performing poorly during conventional paper-and-pencil tests, and preferring to learn knowledge related to real life instead of abstract knowledge. Alternatively, students excelling at in-class activities showed less interest in additional learning activities and tools, to the extent that they even found these activities and tools unhelpful.

Therefore, we developed a new blogging system by creating a novel learning environment that differs from the conventional class setting. Through the blogging system, students can easily receive or share various teaching materials and multimedia features and browse the Internet. This blogging system provides an informal setting and materials for students with low prior achievements to incorporate formal content with real-life knowledge learned through informal learning and explore their interests in a subject. However, for students with high prior achievements, sharing their life experiences on the blogging system provided no significant benefit to the learning achievements they had already done well in, instead adding additional learning time they did not require.

Furthermore, the developed system supplemented the blogging system with the iTM function to address the poor organization of the blogging system. The iTM function can dynamically present the connection between the topics of
formal learning and informal learning. The supplementation not only provided students with a novel learning experience, but also enabled students with high prior achievements to combine formal content, Internet resources, and blog entries by using the iTM function.

Therefore, to help students effectively combine the formal and informal learning of the Citizenship and Society unit, this study proposed a seamless model, as shown below, to provide teachers and system developers with a new perspective. For the seamless model, students in the high achievement group can use informal learning to facilitate the learning efficiency of formal learning. The iTM function was employed as an assistance tool of innovative technology for these students. Conversely, for students with low achievements, they can learn through informal and innovative methods (e.g., a blogging system) to enhance their interests in formal content and further encourage them to incorporate informal content in the structure of formal knowledge, to create a seamless learning environment (CEDEFOP, 2008).

![Image of seamless model for incorporating formal and informal learning by learners in civic education]

Figure 6. Seamless model for incorporating formal and informal learning by learners in civic education

**Conclusion**

For this study we developed a guide system of blogging embedded with TM technology. The blogging system was employed as an informal learning interface for learners to share their life experiences and news (Woodward & Nanlohy, 2004). In addition, teachers could supplement formal learning with related material. Furthermore, the iTM function was incorporated to enable the formal content and structure to be presented in an organized manner, by using a TM and integrating Internet resources and blog entries. Empirical evidence has shown that the proposed BKNS can facilitate students’ formal learning, provide students with novel experiences of innovative technology applications, and enable students to learn through unique practical experiences.

In addition, this study proposed the seamless model for teachers and system designers to fully bridge formal and informal learning. By using this model, we also identified further areas of concern for combining formal and informal learning. For example, would an informal learning style be expanded to other types such as different social platforms, lighter devices, and collaborative learning? Future studies could test the teaching of other subjects with the seamless model. This will be helpful in attaining a greater understanding of designing curricula and learning activities for easier integration of formal and informal learning.

**Acknowledgments**

We wish to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 99-2511-S-025-002-MY2, NSC 100-2410-H029-043 and NSC 100-2511-S-025-002-MY2.
References


Enriching Formal Language Learning with an Informal Social Component

Giuliana Dettori¹ and Simone Torsani²

¹ITD-CNR, Via De Marini 6, 16149 Genoa, Italy // ²DISCLIC, University of Genoa, Piazza S. Sabina 2, 16123 Genoa, Italy // dettori@itd.cnr.it // s.torsani@gmail.com

*Corresponding author

ABSTRACT

This paper describes an informal component that we added to an online formal language learning environment in order to help the learners reach relevant Internet pages they can freely use to complement their learning activity. Thanks to this facility, each lesson is enriched, at run time, with a number of links automatically retrieved from social bookmarking sites. The learners also have at disposal a micro-evaluation system allowing them to share within the environment their like or dislike of the visited web sites, giving their fellow learners indirect guidance for fruitful web exploitation. Our analysis of the retrieved bookmarks for different kinds of learning contents shows that suitable links are actually retrieved by this facility. Moreover, a pilot experimentation revealed that students actually feel inclined to make use of this facility and appreciate both the proposed links and the possibility to receive advice from their peers by means of the micro-evaluation system. This suggests that the added facility actually constitutes an opportunity of informal learning suitably connected with formal one.

Keywords

Informal learning, Online learning, Social bookmarking, Language learning

Introduction

This paper describes and discusses an informal facility that was added to a formal online environment for language learning, so as to offer learners an opportunity to freely complement learning activities by easily reaching out to external Internet pages related to each activity’s content.

Informal learning has been gaining increasing attention in the past decades, especially in relation with adults, as a crucial element to be valued in order to support competence increase in the workplace (Cross, 2006; Colardyn & Bjornavold, 2004; Marsick & Watkins, 2001) and foster lifelong learning (Vavoula & Sharpless, 2009). Informal learning is usually grounded in experience and can influence both cognition and behavior (Hoekstra et al., 2009). Its free and voluntary nature is well expressed by Cross (2009) by means of a powerful, evocative image: “Informal learning is the unofficial, unscheduled, impromptu way most people learn to do their jobs. (…) Informal learning is like riding a bicycle: the rider chooses the destination and the route. The cyclist can take a detour at a moment’s notice to admire the scenery or help a fellow rider.”

Creating opportunities for informal learning can produce a positive return also on formal education, which informal learning complements by putting into play different kinds of strategies and personal goals (Boekaerts & Minnaerts, 1999), as well as by helping to create a wider perspective on the object of study and contributing to increase learners’ engagement (Cross, 2006). A two-way influence between the formal and informal realms has been spotted by some authors; for instance, Ivanova and Popova (2009) remark that in suitably structured environments “informal learning flows supplement and enrich the formal learning process and … formal learning flows are premises for informal learning to occur”.

In order to create a positive synergy between formal and informal activities, a clear connection among them should be established, yet respecting the lack of compulsion and imposed structure that characterizes (and makes appealing) the informal realm. The literature reports several proposals, mostly relying on social approaches and web technologies, targeted at creating opportunities for informal learning and connecting them with formal education (e.g., Bartlett-Brag, 2008; Breuer et al., 2007; Eshach, 2007; Hodkinson et al., 2003; Lucas & Moreira, 2009; Pettenati & Ranieri, 2006). Also this paper aims to contribute in this respect, by proposing an approach to help students exploit the learning opportunities provided by the web, freely and according to personal tastes yet in connection with their formal learning path.
In the next section we summarize the potential of social networks, and in particular of social bookmarking, as an opportunity for informal learning. Then we describe the language learning platform on which we have been working, together with the newly added informal learning facility based on the use of social bookmarking, highlighting its aims, structure, operation and implementation. Finally, we discuss the validity of our proposal by analyzing an example of use, pointing out how the introduction of the bookmark-based facility helps to integrate formal and informal learning, and spotting which appear to be its potential benefits and limitations.

Potential of the social web for language and informal learning

The Internet represents an important resource for language learners, in that it allows them to get in touch with a huge amount of “real” documents providing a rich variety of examples of language use in contexts that can raise the user’s interest (Dudney, 2007). The development of the so-called Web 2.0 also offers a wealth of possibilities to interact, in a variety of ways and formats, with people around the world, especially in social spaces (Thomas, 2010). Such contacts - with people and socially constructed content - which can be exploited by the learners in quite informal way, deciding day by day what to access and what use to make of it, favor meaning construction, knowledge contextualization and skill practice, and are therefore a rich source of language learning, albeit in a different way than formal instruction.

Thanks to the fast flourishing of social spaces where people, unsolicited and spontaneously, exchange any sort of information, opinions and materials, the web is currently considered the place for informal learning par excellence, that extends to potentially anyone the possibilities of informal learning that were once characteristic of the workplace, that is, learning in unplanned way by getting in touch with the experience and expertise of fellows. As a matter of fact, there are many interesting examples with an evident learning potential, which range from the posting of videos (many of which are expressly targeted at showing how to do something), to public forums where people can ask for advice to tackle any kind of problem (hence indirectly learning from each other’s competence), and up to communities of practice in which a closed group of people deliberately shares professional issues and competence, getting in return cultural, cognitive and motivational support.

A common feature in most of such spaces is the possibility for the visitors to express an opinion on the posted materials in synthetic form, that is, by selecting a like/unlike option or attributing a number of stars. This is a simple way to transform lurkers into participants, revealing their presence, and at the same time brings to the fore what has been defined “the wisdom of the crowd” (Surowiecki, 2005), that is, it lets excellence emerge from participants’ appreciation.

Among the many sharing possibilities offered by the Internet, we decided to rely on social bookmarking because it appears a simple, fast, unobtrusive yet powerful tool suitable to be used in the formal learning situation we desired to enrich. Shared bookmarks represent a natural and powerful extension of web navigation tools which provides a mix of direct advice, given by the bookmark’s proposers, and indirect guidance deriving from the preferences expressed by the other users (Millen et al., 2007). Bookmark collections value the human presence in online communities, in that links here are not discovered by automatic procedures, but searched, recorded, tagged and rated by users for their own benefit and voluntarily made available to other users (Hammond et al., 2005).

Social bookmarking aims to create connections rather than new contents (Berger & Tresler, 2010). This recalls connectivism’s view of learning (Siemens, 2005) according to which “[k]now-how and know-what is being supplemented with know-where (the understanding of where to find knowledge needed).” In this perspective, bookmark collections represent a relevant source of learning.

An integrated system

The formal platform

The CliRe platform, first launched in 2008, is used at the Faculty of Foreign Languages of the University of Genoa (Italy) to complement face-to-face courses offered for five different languages (English, French, German, Spanish, Italian for foreigners), with a total of around one thousand registered users (Torsani, 2012). It was developed based
on a traditional, non-interactive approach and mirrors the organization of typical self-study courses, in which a
learning subject is divided into lessons to be tackled sequentially, each focused on a chosen topic.

In this platform, teachers build learning environments for their courses by creating learning units, each of which
includes a number of compound activities focusing on some topic of grammar (e.g., the past perfect), vocabulary
(e.g., weather forecasting), linguistic functions (e.g., asking for directions), culture and civilization (e.g., the
Highlands in the UK), linguistic skills (e.g., listening comprehension) or a combination of them. Compound
activities are the core element of the entire course structure and are made up of different basic activities, such as
listening an audio file, reading a text, analyzing examples, solving exercises of various types (fill the gap, match,
quiz, etc., but no collaborative tasks). Figure 1 shows the page of a typical unit where the compound activities are
accessed by the users in sequence, by clicking on the icons placed along a curved path. Learning units are sequenced
to form modules. Teachers also have the possibility to endow each activity with keywords describing its contents.

![Figure 1. A unit in the formal language learning environment](image)

The platform is quite neutral with respect to the teaching methodology: It only provides teachers with the tools they
need to construct their courses (i.e., to input activities and to sequence them as desired). All learning contents, the
kind and amount of activities proposed, the choice to use which multimedia (if any) to help convey or construct
meanings, as well as the sequencing and combination of topics are completely decided by each teacher, who
determines, by this mean, a pedagogical approach for each course proposed on the platform and takes care to make
the lessons more or less alive, foster some linguistic skill in particular, stimulate reflection, and so on. A number of
tools (a dictionary, a set of grammar flash cards and web resources) can also be attached to each module by the
teacher, so as to provide learners with some support. Different courses offered on this platform can therefore have
different teaching and learning characteristics, even though they are presented in the same format.

Students are requested to complete all the tasks within a compound activity before passing to the next one, all the
activities of a unit before progressing to the next unit, and all the units of a module before moving to the next
module. They are always allowed, however, to return to any part already completed, in order to rehearse it.

After an initial implementation, the platform has been progressively improved by adding functional components
(e.g., a compound dictionary system) apt to further facilitate and stimulate learning. Moreover, the platform is
currently undergoing a further, deep revision aiming to make it suitable to be used for the activities of a newly
established Linguistic Centre. Besides improving the platform, such developments also have the research aim to
investigate how language learning can be made more effective and pleasant even within the context of a traditional,
self-study approach (Poli & Torsani 2010).

**An opening towards informal learning**

A recent step along the line of improving the language platform is the bookmark facility which is the focus of this
paper. It aims to open the platform to the external world in a focused way, so as to suggest relevant opportunities of
informal learning connected with the users’ formal learning activity. It also aims to support students’ motivation by helping them to overcome the sense of isolation and boredom that Nielson (2011) points out as the cause of a fast decrease of interest in the users of autonomous instruction courses.

This facility performs a search on various social bookmarking websites, based on the descriptive keywords specified by the teachers while designing their courses, and displays a selection of likely related and valuable links at the end of each activity. The website suggestions are not statically included in each activity’s page, but dynamically retrieved each time a student accesses an activity. We chose to repeat resource retrieval at every runtime in order to avoid having a static list of resources for each activity, which would be the case if links were retrieved only once. Looking into social bookmarking instead of directly on the web aims to offer links which have already been filtered by other users, and hence are more likely interesting.

This bookmark facility also includes a micro-evaluation system: Each link listed is followed by two clickable icons to let the users who have visited the link express their like or dislike of it. Each bookmark can be voted only once by each student. The expressed evaluations are stored and summed up so as to be shown at the next use of that unit by any platform user. Bookmarks that receive “dislike” marks are moved down in the list and eventually eliminated. Figure 2 shows part of a list of bookmarks displayed in a compound activity, where the keyword assigned by the teacher was the cultural topic “English proverbs.”

Using any of the links and the micro-evaluation is completely discretionary for the students and does not contribute to the accomplishments required for credit; in other words, missing to use these links does not impede the user to move on to the next activity, unlike the teacher-assigned tasks, which are compulsory in order to proceed along the formal learning path.

The possibility to express a personal evaluation of the bookmarked sites and to see the preferences of the classmates is not only a useful filtering device but also has some relevant educational purposes: to stimulate curiosity; to help the users feel part of a community from which they can obtain useful suggestions; to stimulate the users to look at the visited sites with a critical eye. Curiosity is recognized as a component of motivation according to the model of Malone and Leppers (1987). Also creating a sense of community contributes to raise motivation (Reinders & White, 2011) and to keep interest alive (Nielson, 2011). Finally, expressing an opinion requires the users to pay attention to what they consider valuable and to develop critical abilities; this is recognized to have a positive influence on learning (Berger & Trexler, 2010).

Figure 2. Bookmarks listed at the end of a learning unit

Technical realization

From a technical point of view, this facility was developed through a combination of scripting languages and techniques. We exploited the Jquery Javascript class to make a web page automatically call a script that queries several social bookmarking websites, fetches the returned data and outputs them on the web page (Poli & Torsani, 2012). At the moment, the search is made in three sites which offer a good choice of references interesting for
language learners: Delicious (http://delicious.com/), Digg (http://digg.com/) and xMarks (http://www.xmarks.com/).
A higher number of bookmark sites could also be used. In case less than 15 bookmarks are returned from bookmark sites, a search engine (Yahoo) is also queried.

The system uses the descriptive keywords (which were originally introduced to categorize the activities) to query each considered social bookmarking site, and retrieves the resources as suggested by the site’s users; the collected resources are then sorted and returned to the platform users. Since each site is organized differently, different procedures have been designed in order to retrieve as good as possible results. Some, like Delicious, offer a RSS feed service, hence a public RSS file can be parsed, while in other cases the page’s code needs to be analyzed and the results retrieved by means of data mining procedures. Once a list of resources has been collected, our algorithm sorts them so as to have the most likely suitable ones on the top. This is done through a two-step procedure resulting in a numeric value assigned to each resource of the list. This value is first incremented through a simple match relevancy procedure, and then is incremented or decremented based on previous users' evaluations. Based on this value, the system chooses the first fifteen elements of the list, which are displayed on screen, while the remaining ones are left hidden and saved for possible later use.

The whole retrieval and selection procedure is schematized in Figure 3. The system queries different portals through the activity's keywords (A – the search engine is in grey as it is queried only in case social bookmarking websites return too few results); the retrieved links are temporarily stored (B) and sorted according to a keywords match and previous users evaluations (C). The results are finally sent to the activity's page (D) where the user can evaluate them and her/his vote is recorded for future filtering (E).

![Figure 3. Schema of the technical realization of the bookmark facility](image)

**Evaluating the bookmark facility**

In order to evaluate if the described bookmark facility actually fulfills the proposed aims and can be considered a valuable tool to connect formal and informal learning, we need to check three aspects:

- Does it really retrieve relevant bookmarks which are related to each activity’s content, so as to help realize a seamless integration between formal learning in the online environment and informal learning on the web?
- Do the students feel inclined to use this facility (including the micro-evaluation system), and do they find it useful?
- Does this facility actually contribute to support users’ motivation?

Moreover, we will also comment on the cost-effectiveness of our proposal.
Relevance of the bookmarks retrieved

It is obviously not feasible to check the behavior of our bookmark facility in relation with all possible learning contents. Hence, we carried out an analysis of its outcomes on a sample of cases covering the various types of content-based activities that teachers propose in the considered language learning environment, namely grammatical, lexical and cultural/civilization topics. The gathered data are shown in Table 1, which compares the links obtained by searching the same keywords in 3 cases: (1) with a very diffused and appreciated search engine (Google); (2) with our bookmark facility, at an initial use, before micro-evaluations have been expressed; (3) with our facility, after six users had expressed “likes” and “dislikes” for the visited links. For the sake of simplicity, we limited this comparison to the first ten links retrieved, i.e., the length of one Google result page (which is what users mostly consider in a quick search).

We classified the links obtained in 3 groups: (1) *not pertinent* (N), in case of broken link or content not related with the assigned keywords; (2) *pertinent but not interesting* (P), in case of page with topic related to the keywords but limited or redundant content (e.g., more advertising than text and learning activities); (3) *relevant and interesting* (I), in case of pages with content not only related to the keywords but also worth the time to visit them for at least one of several reasons: richness, accuracy, variety and good organization; feedback on activities’ execution; clear and concise explanations; public discussion on the related topic; games involving language use while putting into play some cognitive processes (because cognitive processes such as selecting, classifying, ordering, reasoning and evaluating information imply linguistic choices and are deemed to be favorable for language learning (Ellis, 2003)).

<table>
<thead>
<tr>
<th>Type of topic</th>
<th>Activity’s title</th>
<th>Keywords used</th>
<th>Retrieved by Google</th>
<th>Retrieved by our tool at first use</th>
<th>Retrieved by our tool after 6 evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>Jobs and professions</td>
<td>Professions, jobs, English</td>
<td>5 1 4</td>
<td>3 1 6</td>
<td>- 2 8</td>
</tr>
<tr>
<td>Grammar</td>
<td>Second conditional</td>
<td>Second Conditional</td>
<td>1 - 9</td>
<td>4 - 6</td>
<td>1 1 8</td>
</tr>
<tr>
<td>Vocabulary / Grammar</td>
<td>Prepositional / Phrasal verbs</td>
<td>Phrasal Verbs, Propositional Verbs</td>
<td>- 6 4</td>
<td>1 3 6</td>
<td>- 2 8</td>
</tr>
<tr>
<td>Culture / Civilization</td>
<td>Proverbs</td>
<td>Proverbs, English</td>
<td>2 5 3</td>
<td>3 3 4</td>
<td>1 1 8</td>
</tr>
</tbody>
</table>

*Note. N = not pertinent; P = pertinent but not interesting; I = pertinent and interesting.*

As it can be seen in Table 1, only in one case, a complex grammatical topic, the number of interesting links retrieved by Google is higher than those of our bookmark facility. This can be explained by the fact that grammar-related keywords are quite standard (e.g., “second conditional” almost unambiguously denotes this very grammar feature) and this facilitates the task of search engines. In all other cases, on the other hand, the number of interesting links retrieved by our bookmark facility is higher than those retrieved by Google, even though not excellent, already at first use. Moreover, the number of interesting sites is always higher than the number of pertinent but not interesting ones, likely because the search is made among links already selected by other people; this is often not the case with the results of Google search.

It must also be noted that the use of the micro-evaluation device improves the number of interesting links displayed by our facility in that the “likes” assigned to a link make it move up in the list, while the “dislikes” make it move down, giving the possibility to yet non-voted (and possibly interesting) links to be proposed to users’ attention and evaluation. Repeated use, therefore, increases the number of interesting links in the top positions of the list, hence improving the performance of our bookmark facility. This is in line with the claim of Hammond and colleagues (2005) that social spaces with micro-evaluation facilities progressively increase their worth: “the more they are used, the more value accrues to the system itself and therefore to all who participate in it.”

A list of only interesting links is never reached in the analyzed cases possibly because the varied and ever changing nature of internet makes it difficult to completely eliminate broken links and disguised commercials even from social bookmarking sites. Moreover, only a limited number of micro-evaluations was carried out in this example, and we
can expect the number of interesting links to increase further when the facility is used by a high number of students. This possibility to enhance the outcomes even after a limited number of uses offers teachers the possibility to improve from the beginning the quality of the links proposed by rating a few times the bookmark list before opening the service to their students. This appears to be a good practice in order not to discourage the users of new activities (whose bookmarks have not been rated yet) from carrying on individual web explorations.

In the considered examples, the choice of keywords used was minimal. Analogously to direct web search, keywords are crucial to determine the suitability of the links obtained. Very rich sets of keywords usually lead to poor outcomes, in that it may be difficult to find sites that concern all of them. Also the language in which the keywords are specified has an influence on the outcomes, because it is not uncommon to find pages on a given topic in different languages. In this example, we have specified the target language among the keywords only in two cases (Jobs and Proverbs) for which pages in different languages would very likely be found otherwise. In the other two cases, the strong specificity of the grammatical expressions involved (Second conditional and Phrasal verbs) was sufficient to obtain links to English pages.

**Interest and appreciation of the students**

We run a pilot experimentation in order to test students’ appreciation and actual use of this facility. To this end, we prepared one module of an English course at intermediate level, including all the various kinds of activities supported on the platform, and made a call for voluntary users. We intentionally avoided involving a class because we wanted our experimental subjects not to know each other and not to be in condition to influence each other’s behavior and perception by communicating outside the platform. Sixteen students (ten females and six males) were enrolled in this pilot course, with different backgrounds and learning needs, of age between 20 and 42. None of them had used the CliRe platform before the introduction of the bookmark facility. All the participants were sent access credentials at the same time and were asked to complete their activity in the pilot course within three weeks.

We decided to make a qualitative analysis of the experience’s outcomes because we thought that a rich description of participants’ actions and perceptions could help to shed light on the tool’s actual functioning and usefulness more than analyzing a quantification of visited websites and related micro-evaluations. We collected the participants’ stories and opinions by means of semi-structured interviews so as to make sure to touch the same points with all of them yet leaving them free to personalize their reports. We asked: (1) if they had used the bookmark facility and how extensively; (2) how they had chosen the links to visit (e.g., links with a good evaluation or with an appealing title); (3) how they judged the visited bookmarks as concerns pertinence to the formal activity just carried out, interest and accuracy of the content, and usefulness for their learning; (4) if the possibility to reach related websites from within the very course platform had been useful, made the course more interesting and rich, and helped them to learn more. We also asked if they were familiar with social spaces and with the micro-evaluation system.

Among the collected interviews, we selected two which appear particularly interesting in that they represent two extreme positions, that is, the most and the least enthusiast participants. Their reports are summarized below.

E. is a working adult aged 42, who has decided to improve her English because she has realized that it is necessary in many occasions in the current globalized world. She is rather busy with her job (she works in a day care) and family commitments, so she was happy to have an opportunity to take a course online, even though of limited extent. She was not member of any social space, but she was aware of the existence of micro-evaluation because she had seen (but never used) it on some public web sites like YouTube. In our online course, she made use of the bookmark facility wherever she found it, and deeply appreciated this possibility because at the end of the formal activities she always felt the need of some more practice to consolidate what she had been trying to learn. She explored only two or three links for each activity (because she preferred to explore few in depth rather than many superficially), always selecting the links among those who had already received positive evaluations by some course mates. She explained that she did not want to spend time exploring new websites; in the past, she had already tried to learn English by looking for web resources but with scarce success because she had usually met sets of exercises of scarce content and limited interest or commercial websites trying to sell (expensive) English courses, so she had given up with trying to learn English on the web, and she would certainly not decide to look there for more materials without this facility within course. On the other hand, she was enthusiastic of the web sites she had reached by following these bookmarks, because of the richness, variety and clear organization of their content, so much that she had saved many
of them to keep using those resources after the end of our experiment. Among the sites visited, she mentioned with particular appreciation a page from Yahoo Answers with a synthetic and effective explanation of a grammatical topic that she found particularly difficult (the second conditional). After visiting the web sites, she always used the micro-evaluation system to express her appreciation, and commented in this respect that she had appreciated very much to use good suggestions from other users, and hence was glad to reciprocate by adding hers. Her overall judgment of the bookmark facility was that “it amplifies the learning space by means of nice possibilities.”

M., on the other hand, is a graduate student in Natural Science aged 23 who wants to keep updated and possibly improve her English knowledge because she is aware that it is useful for study and future profession. She is familiar with social spaces, having been on Facebook for a while, but never felt very fond of them. In our online course, she visited a few websites for each activity, usually between three and five, choosing them based on their titles, avoiding YouTube pages, that she considers mainly distracting, and disregarding her course mates’ evaluations; in fact, she said that she often gave preference to sites not yet voted by anybody, so as to contribute some more evaluations to the bookmark list. She found sites of very different quality, some really interesting, others scarcely pleasant because of an excessive amount of publicity banners, and also some broken links. The deception for a few of the links visited limited her enthusiasm for the course in general and for the bookmark facility in particular. This did not prevent her, however, from completing the course and she kept visiting suggested bookmarks; she also diligently expressed her opinion with the micro-evaluation device for all sites visited. Her overall judgment was not completely negative, however, in that she admitted that the presence of external links contributes to support her interest, though this varied in dependence on the quality of the links visited. Also M., like E., reported that she would not go to the web and look for sites related to the activities’ content if she had not been prompted by the presence of the bookmark lists.

The accounts of these two students confirm what we reported above concerning the relevance of the bookmarks retrieved by our search procedure, that is, that for every activity several pertinent and interesting ones were always retrieved, together with some irrelevant ones. The different levels of satisfaction of the two students largely depend on their different ways to select the websites to visit: E. played it safe, selecting only links already valued by someone, and this explains her enthusiasm for the materials retrieved. M., on the other hand, limiting her choice to web sites not yet voted by any course mate, took the risk to find unsuitable connections. The fact that she, nevertheless, reported to have found several very interesting ones marks a point in favor of the amount of good links retrieved by our procedure and of the possibility to increase the number of pertinent and interesting links over repeated use of the micro-evaluation system. This underlines the potential of the proposed facility to foster informal learning in connection with formal one. We can add, making reference also to the behavior of the other students involved in this experience, that in average users tend to chose a mix of positively evaluated and yet to be voted links; such behavior has the advantage for the user to grant a few very likely interesting sites, and at the same time contributes to improve the overall quality of the displayed links.

Both students visited a limited number of links with respect to those listed by the facility, a behavior which was common with that of all the other students involved in the experience. This is not surprising, since visiting web sites is rather time consuming, especially if they are rich and compound, and also because not all sites may be of interest of every user, independently of their pertinence and intrinsic interest. Listing more sites than the students usually visit, but not so many to make the selection difficult, appears a positive feature to grant the students more freedom of choice, and marks a point in favor of the actual informality of the described bookmark facility.

We are aware that an experience like the described one, which is limited as for number of participants and temporal extension, cannot answer all possible questions concerning the value of the proposed facility. Further investigations need to be carried out to check if the users’ interest for it remains high over longer periods, and how widely it can vary based on users’ characteristics. The positive outcomes of this experience, however, show that students’ appreciation and actual use is possible, and hence this approach to the integration of formal and informal learning is worth being pursued.

**Support to motivation**

In order to analyze the potential support to motivation, we need to see if at least the two motivation-related features mentioned above can be found in the accounts of the students involved. This means checking if students’ curiosity was stimulated and if they felt part of a community of users, even though working in isolation.
In both cases, we can find affirmative hints. Curiosity is suggested by the fact that both students kept exploring external links throughout the course, and in particular in the case of M., who continued doing so despite the number of unsatisfactory links found. The awareness of being part of a community of users is underlined by the fact that both students paid attention to the votes already expressed by the other users, even though they decided to make a different use of this information. They also felt the wish to always contribute their own evaluation, as explicitly specified in both interviews. These aspects were always present, in smaller or bigger measure, also in the interviews of the other students, who usually admitted not to feel inclined to complement class activities by freely searching related materials on the web, but to have been attracted to visit a few of the proposed links either by noticing an appealing title or by the fact that it had received positive peers’ evaluations and this induced them to see by themselves why.

We need to point out, however, that making an overall evaluation of the facility’s positive impact on users’ motivation is a complex task which can hardly be carried out in a short experience, because motivation variations can meaningfully be measured only over longer periods of time.

**Cost-effectiveness of the proposed facility**

A positive point of this bookmark facility is its simplicity. Its activation is not a burden for the teachers, who need to enter proper keywords for each compound activity only once, and are not required to engage in more time-consuming tasks, like selecting a set of meaningful extra resources to suggest, and maintain it updated. Neither is it a burden for the students, as no compulsory extra workload is added: the students are free to follow, at their own pace, only the suggestions they deem appealing, and even to save them for later use. The tool’s run time, moreover, is reasonable, so that, by the time the users have completed the activity’s tasks, the bookmark list is ready to be displayed without any extra waiting.

Some periodical maintenance work, however, is required of the programmer, because data mining depends on the structure of the source pages analyzed, which often change from time to time.

**Conclusions**

This paper describes a possible approach to support seamless integration of formal and informal learning, which we developed for the users of a formal language learning platform. The idea of using social bookmarking in education is certainly not new; what is new in our proposal is the way we exploit it to encourage individual web exploration.

Our approach consists in extracting a number of suggestions from social bookmarking sites based on keywords characterizing each activity’s content, and to select among them what appears more suitable to the activity’s learning content. The retrieval and selection procedures that we developed for this purpose are quite general and do not limit the application of this approach to language learning, but lend themselves to be applied as well for other subjects.

The presence of a list of links to external resources at the end of each formal activity stimulates the students to extend their learning to useful sites outside the formal learning platform. The use of already published bookmarks and the local micro-evaluation device guide such voluntary web exploration towards resources related to the formal learning contents and at the same time to the user’s interests, hence helping to establish a smooth continuity between formal and informal learning activities.

Several aspects of the described facility underline its informal nature: The proposed external links are not teacher-defined but originate from social bookmarking websites, where they have been voluntarily suggested and rated by other users; visiting what links (if any), and expressing individual appreciations is not compulsory but completely up to the students; the number of links is much higher than students usually like to visit, hence granting them space for choice based on individual decision. All this actually recalls Cross’ metaphor of informal learning (2009) cited in the introduction, in which the informal learner is likened to a bicycle rider who is free to individually decide route, detours and destination.
References


PKS: An Ontology-based Learning Construct for Lifelong Learners

Flavio Manganello*, Carla Falsetti, Luca Spalazzi and Tommaso Leo

Department of Information Engineering, University Politecnica delle Marche, Italy // f.manganello@univpm.it // c.falsetti@univpm.it // spalazzi@dii.univpm.it // tommaso.leo@univpm.it

*Corresponding author

ABSTRACT

This paper addresses adult lifelong learners, i.e., persons interested in learning or compelled to learn during their working life but not able to, or not interested in participating in formal learning. These learners are motivated and self-aware enough to self-direct their learning, are presumed to be novices with respect to the needed knowledge and have a limited technological uptake. This paper will outline their main differences from regular learners, in particular, for the use of social media to improve learning skills. The reference pedagogical model is andragogy. A Service-Oriented Architecture, named Personal Knowledge Space (PKS), is proposed to support our lifelong learners in selecting, organizing, and retrieving information; in streamlining interaction processes among learners, services and resources and finally in empowering control and trust of personal relationships born during the learning processes. The PKS is mainly based on the exploitation of semantic tools (ontologies) and web services. Use cases describing the PKS architecture and one scenario of PKS use are presented.

Keywords

Informal learning, Andragogy, Semantic web, Service oriented architecture, Professional development

Introduction

This paper focuses on adult lifelong learners (LLLS), i.e., persons interested/compelled in learning during their professional life but not able/interested to learn in educational institutions. Their specific learning characteristics can be summarized as follows:

- the learning process is intentional and self-motivated;
- learning pace and time are in charge of the learner;
- learning objectives are chosen by the learners, usually refer to knowledge domains not yet mastered and are mainly related to building usable knowledge relevant to professional purposes;
- the attainment of the learning objectives doesn’t lead to a certification but is appraised by the learners themselves;
- knowledge and learning resources for the learning process are searched for by the learner;
- the learner doesn’t ask educational institutions to provide structured support, but may want to consult learning resources offered by institutions (i.e., handbooks, courseware, …).

Therefore, the learning process of our LLLs may utilize formal, non-formal and informal learning resources for professional development. They would require the professional to interact with experts or peers in the targeted domains. This learning process encompasses most of the ten attributes of informal learning described in Boekaerts & Minnaert (1999). Motivation and ability to self-regulate the learning process, are recognizable features in many professionals.

Anderson (2005) and Klamma et al. (2007) argue that web 2.0 (Oreilly, 2007) and social software may support distance and lifelong learning. Broadly speaking, social software applications are “communication tools and interactive tools often based on the Internet” (“Social software,” n.d.). Web 2.0 technologies may be used to support the development of LLL skills (Dunlap & Lowenthal, 2011). However, they require the learner to master technological skills. Our LLLs might not have such skills since in most cases they are “digital immigrants” (Prensky, 2001). They might be exposed to information overload (Sweller, 1988) and to the syndrome of “being lost in social space” (Dron & Anderson, 2009).

To enlighten the characters of our target LLLs, let us mention some characters of the regular students within educational institutions. They are usually “digital natives” (Prensky, 2001), but frequently lack the competence to exploit web 2.0 technologies for learning purposes, and the ability to self-regulating their learning. Training is needed to fully exploit the learning power of these technologies.
Educational institutions are committed to offer their students tools to assist them in carrying out informal learning in support of formal learning (Dron & Anderson, 2009). For this purpose, Personal Learning Environments (PLEs) seem to be a promising approach. The current PLE implementation within educational institutions (Dabbagh & Kitsantas, 2012; Klamma et al., 2007; Pettenati et al., 2007) and professional organizations (Cheng et al., 2011; Scherp et al., 2009) seems to be characterized as follows:

- the knowledge domain and its specific language are usually defined by a teacher or a tutor;
- the students define their personal profile, usually only in social terms;
- the learning needs are defined with the teacher’s support;
- after the profile’s definition, the students begin to shape their PLE;
- the learning objectives are defined by the institution/organization, not by the students.

On the contrary, our target LLLs should specify by themselves their learning profiles in terms of domain, objectives and social actors to be found, as well as master the knowledge domain language.

Therefore, our objective is to propose a model architecture able to assist our target LLLs in finding, reordering, classifying, interacting with the proper knowledge resources, and sharing them with other persons having similar requirements. Such an architecture could assist in the birth of some “semiotic social space” (Gee, 2005) able to support a community of learners in the process of sharing and creating knowledge.

The non-homogeneous level of technology uptake has to be taken into account, together with the need of personalizing the learning experiences. This implies a number of relevant issues: proper profiling of the LLL, semantic exploration of the resources available on the web, identification of candidate persons to enter in touch with. Moreover, support for suitable tagging of such resources should be provided.

We would name this architecture “Personal Knowledge Space” (PKS). The research objective of this paper is to demonstrate that such architecture can be implemented by using some of the currently available web 2.0 tools and services.

Related work

We will consider some skills that regular students are trained to learn at educational institutions. These skills refer to the ability of exploiting web 2.0 technologies and self regulating the learning process. In particular, we will consider research relative to the PLE use at educational institutions. The purpose is to identify technology functions able to replace the guidance given by the institution (teachers, tutors, curricula), at least partially. Of course, this survey will be partial because of the richness of the relevant literature.

There are several definitions of PLE. According to Attwell (2007), PLE is not an application but rather a construct describing the use of new technologies for learning. In this paper, we will refer to PLE as a metaphor generating at least two research mainstreams:

- the educational mainstream focuses on pedagogical and instructional issues related to the adoption of PLEs, also within formal contexts (Häkkinen & Hämäläinen, 2011; Pata, 2009). PLE helps learners to take control of their own learning (van Harmelen, 2006).
- the technological mainstream mainly investigates components, tools, or applications suitable for integration into PLEs (Chatti et al., 2009; Drachsler et al., 2009; Wilson et al., 2009). PLE is described as a mash-up application.

As for the educational mainstream, notice that the ability of self regulating the learning process is a fundamental feature of a LLL (Boekaerts, 1999). As shown by psychological research, such ability dynamically changes due to motivation. Motivation is heavily influenced by psychological and social aspects involved in social software use, as clarified by Pereira et al. (2010). Their model fits well with the concept of “semiotic social space” (Gee, 2005) and with the concept of “socio digitization” (Latham & Sassen, 2005).

As for technology, the learner can create her/his own learning mash-up that leverages components and contents generated by learning service providers and other learners around the web. So a PLE may be seen as a learner-centered and learner-designed content aggregation and connections management tool. The focus is on creating knowledge through reification (Wenger et al., 2002).
It can be argued that the features of the available learning environment may enforce motivation or foster disaffection in the development of the learning process. According to Conole (2010) another key issue is that the technology uptake of many potentially effective LLLs is not sufficient to enable them to fully exploit the web 2.0 opportunities. Notice that, as shown by Valtonen et al. (2011), there is no real differences between “digital natives” and “digital immigrants.” Therefore, it seems reasonable to consider such issues as specifications for usability and accessibility of social software.

The educational institutions are committed to offer their students tools to assist them in developing informal learning patterns in support of formal learning. Some proposals related with the concept of PLE are driven by the concern to give students the skills needed to become LLLs. Pettenati et al. (2007) and Dabbagh & Kitsantas (2012) highlight that students need specific training to empower their sense of personal agency.

In our opinion the key point is the intention to train students to become knowledge prosumers (producers and consumers in parallel) able to exploit the interoperability between content and services in a ubiquitous environment. Klamma et al. (2007) highlight the objective of training students to participate in the creation of web 2.0 by developing activities through social software tools and environments. In corporate settings too, some professional organizations aim at exploiting web 2.0 communities to create, organize, share, and utilize enterprise-relevant knowledge (Cheng et al., 2011; Scherp et al., 2009).

On balance, PLEs used within institutional settings seem to have the following limitations:

- the start up of the PLE is proposed and/or guided by the educational institutions (Conole, 2010; Drachsler et al., 2009) and the teacher has to act as tutor or facilitator;
- technological skills are needed to manage the mash-up to integrate different services;
- in the student profile, knowledge needs/objectives are implicitly defined by the curriculum offered by the educational institution.

More limitations and/or disadvantages are presented by Ivanova (2011). The architecture proposed in our work is aimed at reducing or eliminating some of these limitations.

The enabling technologies for PLE rely on the Software-as-a-Service (SaaS) delivery model. In such a model the software owner is separated from the user (Laplante et al., 2003; Turner et al., 2003). In the last decade, the SaaS model has become one of the most common delivery model in several application domains: among them, learning (Learning-as-a-Service or LaaS) and social networks (Social SaaS) are related to our work. The LaaS model (Spaniol et al., 2008) aims at providing learners with the most suitable contents anytime/anywhere according to the SaaS model. Furthermore, the Service Oriented Architecture (SOA) implicit in the SaaS model makes the interoperability of resources easier. Notice that the most popular social networks (e.g., Facebook, LinkedIn, and Google) adopt a SaaS delivery model.

Apparently, interoperability among heterogeneous services, which can be difficult or even impossible to attain when there are different schemas and different names for the same concepts, is not systematically tackled by PLE research. Semantic tools could help overcome this problem. Recent researches suggest exploiting social resources for defining ontology specifications for PLE (Ivanova, 2011; Ivanova & Chatti, 2010). In our view, a viable solution is represented by the Semantic Web (Berners-Lee et al., 2001) and the Social Semantic Web (Torniai et al., 2008), where services and data must have access to structured collections of concepts and relations between concepts (i.e., ontologies), and sets of inference rules. Concepts are usually thought of as sets of objects or individuals whereas rules can be used for automated reasoning (Baader et al., 2003).

Most authors agree that an ontology should be defined in a formal language suitable for automating reasoning. Well known languages are:

- OWL (Motik et al., 2005), based on the Description Logics SHIQ(D) and SHIOQ(D) (Sirin et al, 2007);
- RDF (W3C, 2004), based on multigraphs;

An ontology should be a unique, universally accepted formalization of the domain. This is indeed a strong assumption. However, recent trends show the definition of a large number of ontologies related to many domains.
Several companies and organizations are defining, or have already defined, standard ontologies for specific domains (Bodenreider, 2004; Nayar & Beldona, 2010). We will assume, in the rest of the paper, that it is possible to use standard ontologies for a given domain.

PKS specifications

Since the PKS conceptual model is strictly related to learning and knowledge management, we need to scaffold technical requirements within the suitable pedagogical framework.

Pedagogical model

The PKS is based on social-constructivist learning theories. The main references are constructivism and social-constructivism (Jonassen & Land, 2000; Varisco, 2002). Other theoretical references are networked learning (Blackall, 2005) and connectivism (Siemens, 2004). Within this theoretical frame, the concept of community is central.

The present paper mainly refers to “andragogy” (Knowles, 1990). Even if criticized by some authors (Merriam et al., 2007), it seems that the characters of adult learning require the definition of specific educational frameworks: The learning path must respond as accurately as possible to the needs of the recipient; for learning to be effective adults need to be actively involved in a learning process consistent with her/his specific needs; moreover, learning must take place in a collaborative environment, where learner can interact with other learners or with persons she/he recognizes as experts.

In accordance with this pedagogical model, the PKS would help the learner in:

- selecting, organizing, retrieving information, and reducing the information overload;
- streamlining the interaction processes among learners, services and resources, with benefits on the “lost in social space” phenomenon;
- empowering control and trust of the personal relationships.

PKS technical and functional requirements

Our target LLL is a motivated adult learner, self-aware enough to self direct their own learning process in order to acquire usable knowledge. Such knowledge is frequently multi-domain and to some extent tacit (owned by experts to be identified and/or to be built up through social interaction with peers). The relevant formalized knowledge has also to be identified and collected.

Our target LLL is initially a novice, has a limited technological uptake, and has to deal with documental and human resources (see Figure 1). He has basically three main tasks to fulfill:

1. to retrieve resources from the web space ( “Semantic Searching,” Figure 1);
2. to organize and/or classify them, (“Semantic Collaborative Tagging,” Figure 1);
3. to “use” them (“Contacting People” and “Accessing Documents,” Figure 1).

From a knowledge management perspective, the PKS is for the learner a personal environment capable of representing and managing possible states of her/his knowledge within one or more domains of knowledge. Such domains of knowledge represent concepts known and knowable by the learner (Doignon & Falmagne, 1998). Therefore, resource retrieval and resource organization have to be driven by a module devoted to the construction and the dynamic update of the learner profile (“My Profile,” Figure 1). Furthermore, each learner may produce documents to be shared with other learners in the PKS network (“My Docs,” Figure 1). Such documents must be organized and classified.

Due to the large variability of the learners’ characters and requirements and of their knowledge contexts, we propose semantics to allow unambiguous communication among human actors and digital resources. The proper semantic tools are ontologies. Databases and taxonomies do not have the feature of being universally agreed upon. The initial state and parameters of the learner profile should be set up by the learner on the basis of suitable ontologies. While
learning progresses, adaptivity (Brusilovsky, 2001) and personalization of the learner profile are to be managed at a semantic level.

Finally, all the PKS functionalities have to be invoked by means of a friendly user interface built up as an intuitive and easy to use Web page.
**PKS architecture**

The PKS architecture consists of one module for each main use case. Each module deploys three basic services: searching for resources; organizing and classifying resources; and “using” resources (contacting people and accessing resources). Furthermore, to support a population of LLLs and to foster the creation of specific communities, PKS has been designed with a SOA adopting a SaaS delivery model. The architecture is depicted in Figure 2. The server is depicted by means of a UML (Unified Modelling Language) (Hunt, 2000) component diagram to make the deployed services evident. The server has also a repository (the “PKS Repository” in Figure 2) containing the profile and all the retrieved resources of each LLL. The client (PKS Desktop) is depicted by means of a workflow diagram using the Business Process Management Notation (BPMN) (White & Miers, 2008) to make evident the activity sequence adopted by the learner for her/his learning process. The client allows the learner to perform the operations represented by the four main use cases. The client has also a “Local Repository” with learner own resources and profile. The learner can synchronize her/his local repository with the PKS repository. Finally, Figure 2 makes evident that the PKS server is interfaced with social software and other web resources (e.g., LinkedIn, Google, repositories, and so on).

A SOA has to be preferred to a peer-to-peer architecture. Indeed, in a P2P architecture resources are only local, even if shared with other users. Therefore, a resource becomes temporary unavailable when the owner is not connected. Furthermore, our target LLL might want to access to her/his resources from different devices in an ubiquitous computing fashion. The SOA solution allows the learner to have access to the same set of resources independently on the device she/he is using. Finally, the SOA is interoperable with other tools.

**Use case: “My Profile”**

The profile consists of a set of attributes describing a learner; there is one profile for each PKS user. Useful attributes for a basic learner’s profiling can be found in Brusilovsky (2001) and Wilson (2005). Here we consider the following categories coherent with the Europass Model for CV (Karampiperis et al., 2006; CEDEFOP, 2012):

1. Personal information;
2. Desired employment / Occupational field;
3. Work experience;
4. Education and Training;
5. Language(s);
6. Personal skills and competences.

The learner will describe: under category 2, the current field of interest; under category 3, the held competences; under category 4, the certified past knowledge, such as degrees; under category 5, the known languages (to filter the retrieved resources); under category 6, the social networks she/he participates in and the relative accounts, and finally the level of ICT uptake.

The profile has an initial default configuration (Figure 3).

This manner of building up the profile would avoid the risk of inconsistency and ambiguity connected with a completely free bottom-up approach, due to the different language and cultural background of each LLL.

To properly define the profile of the PKS user attains the identity issue. In the context of Knowledge Management this issue has to be dynamically related to the current topic of interest for the PKS user.

To be effective the PKS interface has to assist the learner in self-defining and to specify the present profile in coherence with the profile specifications of the existing social networks he may want to join.

On the basis of the information provided by the learner under categories 2 and 4, the PKS would propose one or more “knowledge contexts”. Those are presented to the learner by means of a tabbed document interface. Tabs allow to visualize separated, different, and contextual interfaces, which only present meaningful information on a particular knowledge context. Each tab is composed by several graphical elements (Figure 4).
The PKS asks the learner to type in a description of the knowledge context and some keywords and then automatically looks up for corresponding ontology(ies) in its repository. If no suitable ontology is available, the learner can import a known ontology, and/or ask the PKS to search for Owl ontologies in Google. If nothing is found, the PKS suggests using WordNet (Miller, 1995). In this case, the learner would use WordNet concepts similar to the original keywords. Next, the PKS presents a tree corresponding to the ontology, or a vocabulary of terms corresponding to WordNet.
After the user has selected meaningful concepts, the PKS will start the search of documental and human resources. This manner of setting the semantic search is again aimed at avoiding the risk of inconsistency and ambiguity. The use case is implemented by the My Profile module that provides the following three basic services: Insert Profile, Import Profile, and Import Ontology:

- the Insert Profile service consists in guiding the learner step-by-step in inserting a profile according to Europass;
- the Import Profile service imports relevant information about the learner from the profiles inserted in her/his accounts at other social software;
- the Import Ontology service selects the domain ontologies related to learner’s objectives, interests, previous knowledge, and skills. The PKS proposes the ontologies containing concepts related to the keywords specified in the knowledge context tab interface (Figure 4).

The ontologies here considered are OWL ontologies, which are based on description logics.

Use case: “Documental Resources”

Documental resources are documents (in different formats), web sites, multimedia resources and other learning resources available online. This use case is implemented by the Documental Resource Management module that provides the following basic services: Semantic Search, Semantic Tagging, and Document Access.

The Semantic Search service formulates a query and retrieves a set of instances answering the query, within and outside the PKS. The PKS provides the learner with an interface to navigate the selected ontology and to extract concepts and attributes to be used for internal search. The same concepts and attributes can be used as keywords for external search. In the case of internal search, the selected concepts and attributes are translated in standard SPARQL-DL queries (Sirin & Parsia, 2007) by a PKS procedure and answered by Pellet OWL Reasoner (Sirin, Parsia, & Cuenca, 2007) working inside the PKS server. In the case of external search (semantic/non semantic), the keywords are used to search for more resources by means of popular search engines (e.g., Google) and/or repositories of learning resources (e.g., Merlot, MIT Open Courseware, …). So, external (semantic and non-semantic) search engines are mashed-up.

All the results are presented to the learner in a single interface (Figure 5), allowing to select whatever is deemed useful.

If the retrieved resource comprehends standard metadata, they are automatically parsed and used to classify the resource. The following elements from the DCMI (Dublin Core Metadata Initiative) are parsed: title, subject, description, rights, date, language, rights’ holder, and identifier. This may be the case of a learning resource retrieved from an institutional repository, or an open repository such as Merlot (http://www.merlot.org). If the retrieved resource is not indexed by metadata, the PKS user can tag it by using the concepts and attributes which have generated the semantic query. Moreover, after reflection, personal tags can be added by means of the Semantic Tagging service.

This service aims at tagging retrieved resources in order to allow for further retrieval. The learner can therefore tag a resource with a set of assertions representing relevant concepts and attributes. Resources coming from PKS Repository have already been tagged by other users, but the learner can decide to add some more tags for her/his private use or for sharing with other learners. So the task becomes a sort of collaborative tagging (Marchetti & Tesconi, 2007). Tagging a resource is made through a proper interface (Figure 6) and consists in selecting concepts and attributes from the ontology. Furthermore, the PKS suggests the most popular tags. Other suggested tags correspond to the elements “subject” and “description” of the DCMI. Therefore the PKS Repository is a semantic knowledge base. From a technical point of view, the set of tagged resources is the assertional box of the knowledge base whereas the ontology is its terminological box (Baader et al., 2003).

The Document Access service allows the learner to access a selected resource. This service is interfaced with the most popular resource readers/players such as PDF, DOC readers, and so on. Some resources may be accessible only to authorized users. For example, resources delivered by a publisher or by an institution. Therefore, this service must be able to deal with several learner’s accounts. For the experiment presented in this work, we used Google and Google Scholar where account is not needed.
Use case: “Human Resources” and use case: “My Docs”

These use cases are implemented respectively by the Human Resource Management and by the My Docs modules. The first one provides: Semantic Search, Semantic Tagging, (both similar to the corresponding services of Documental resource management module) and Contact services. These services allow the learner to communicate with people selected as experts thanks to Semantic Search. This service must be interfaced with some popular social networks. Therefore it has to deal with several learner’s accounts. For the experiment presented in this work, we used LinkedIn. If the user has already an account on the selected social network, the PKS will use the information provided in the profile to automatically authenticate the learner; otherwise, learners will need to create the proper account and then update their profile.
The second module provides *Insert Document*, and *Semantic Tagging* services. The *Insert Document* service aims at inserting in the PKS Repository a document or a resource directly produced by the learner. This service is able to work with different file formats.

The *Semantic Tagging* service allows the learner to tag the resource with a set of assertions in a collaborative fashion, as in *Documental Resource Management* modules.

**Scenario of use**

A given LLL begins to use the PKS and needs resources in a given domain, for instance in the case of professional development of Complex System Professionals (Leo et al., 2009). Here we consider an Italian junior systems engineer that has to start working in the field of renewable energies and domotics, not previously studied.

The activities of such engineer through and within the PKS are as follows:

**My Profile**

*Insert Profile.* The learner is guided step-by-step by the PKS in filling up her profile according to the Europass CV model (see use case “My Profile”). She defines: energy resources, domotic, domotics as field of interest; control system design as work experiences; engineering degree as education and training; Italian and English as languages; medium ICT uptake and some social network (LinkedIn) as personal skills and abilities.

*Import ontology.* In our case the PKS does not have related ontologies in the Repository and starts searching among Google.owl ontologies using the keyword “filetype:owl energy resources, domotic, domotics.” The result shows (at November 2011):

- a) https://www.auto.tuwien.ac.at/downloads/ThinkHome/ontology/0_87/EnergyResourceOntology.owl;
- b) http://amigo.gforge.inria.fr/owl/Domotics.owl;
- c) http://elite.polito.it/files/releases/dog/dogont/DogOnt-1.0.5-MaisonEquipee/DOGOnt.owl.

The PKS shows the contents and structure of the ontologies using the interface depicted in Figure 4. The LLL decides to adopt the first ontology in the list because it covers both non formal categories (renewable energy) and disciplinary domain (domotics).

At this point, the learner can search for documental resources and/or human resources.

**Documental Resources**

*Semantic search.* The learner navigates the ontology and selects concepts and attributes. In this case, she selects the concepts “Energy Provider” and “BTicino Component” (which has “Domotic Network Component” as superconcept); they are used to search for documental resources both inside and outside the PKS. In the present case, no documents are available in the PKS repository, so she starts searching Google by keywords. One of the search key selected is “Energy provider,” refined by the learner into “Energy provider Marche region Italy”; among the results, one document of interest is a Wikipedia page about Hera, a company operating in the energy market. Another search key is “Domotic Network Component BTicino Component”. With this second key, several results are listed. Among those, our LLL is interested in the “Social network optimization: the case of BTicino” document.

*Document access.* Firstly, our LLL explores Wikipedia page relative to Hera and consult the page, containing number of interesting links. She decides to save the page. Secondly, she explores and downloads the “Social network optimization: the case of BTicino” document authored by persons apparently authoritative or already known to her; so these are contacts that she will refer to using the Human Resource Module.

*Semantic tagging.* All the downloaded documents are semantically annotated, put in the PKS client and then synchronized with the PKS server. Semantic annotation is guided by the procedure described under *Semantic
tagging. In our case the Hera company web site is tagged with “Energy Provider Marche region Italy”; the “Social network optimization: The case of BTicino” document is tagged with “Social Network” and “BTicino.”

**Human resources.** Starting from the authors of the “Social network optimization: the case of BTicino” document, our LLL uses her PKS client to handle the authentication for accessing social networks and using their searching services. In the prototype implemented for this work, LinkedIn has been used. The profile of the expert in “project management” is retrieved from LinkedIn and then the LLL has the opportunity to use the social network to contact him.

Our LLL is also interested in searching designers working in the domain. In the Advanced search module of LinkedIn, the PKS automatically proposes the field Keywords (desired employment, as described in the profile), Title (the highest degree of Education and Training) and Industry (non formal category in the interests Renewable energy).

In this last kind of research, the search driven by PKS has to be refined in various ways by the LLL, particularly if the search is not limited to her own country. The translation in different languages of the main search key can produce improper results, due to the different cultures of different countries; there is a problem of semantic inconsistency of the syntax of the research key.

**Semantic tagging.** The profile of the retrieved human resource has to be semantically annotated. The default tags are those published, on LinkedIn profile, and the relative rating, if available. In our case, tags and ratings are “Skills” and “Recommendations” from LinkedIn.

**Discussion and conclusions**

A SOA has been proposed to support knowledge building and sharing for the professional advancement of adult LLLs, not able to attend institutional courses for their learning goals. The distinctive characters of such learners are motivation, capability of self directed learning, and limited technology uptake. Such population represents a large set of professionals, having a traditional educational background and compelled to face new professional challenges in domains they are partially or totally unaware of. They frequently need to discover and acquire tacit knowledge relevant to their professional needs and make it explicit.

The PKS architecture has a number of similarities and some fundamental differences with the PLE construct. One evident difference is that there are neither teachers available nor institutional tutoring/mentoring to plan the learning process and suggest experts. Therefore we suggest a proactive architecture, using semantics, within a collaborative environment. From the functional point of view, the key element is the construction of the user profile based on a specific ontology and the search of ontologies relative to the LLL domain of interest. The LLL has to choose among the ontologies proposed by the PKS.

This is a key point: Indeed the LLL is initially a novice in the domain classified by the ontology. The chosen ontology defines the semantic space for searching documents and people, and organizing the results. It is not rare the case of multiple ontologies relative to the same domain. Moreover the professional problem addressed by the LLL can be multi-domain. This choice has to be made very carefully and has to be open to modifications with the evolution of the LLL competences.

The PKS provides the LLL with tools for adaptivity in respect to the personal profile and for personalization. This fundamental difference with PLE asks for relying again on semantics to assist the learner in the knowledge resources retrieval and organization and in the definition of their personal identity. This is a crucial point for the LLL to be recognizable in the social space; in this way peers and experts, with which to establish a community of knowledge prosumers in relation to the present object of interest, can be found.

On the technological side, the PKS is part of the SaaS delivery model, in a web 2.0 environment. The choice of a SOA seems apt, in principle, to guarantee system flexibility and adaptivity. The choice of relying onto available ontologies seems, in principle, apt to allow ease of use, reducing the risk of ambiguity and inconsistency both in the search of results and in the definition of personal identity. Actually the present diffusion of ontologies would not be
enough for a wide use of such an architecture, but the widespread interest in these topics will hopefully let their
number increase rapidly.

The presentation of the PKS architecture is a first step. Future work, in our view, will focus on dynamically updating
the profile of the LLL according to the progression of Knowledge construction and collaboration activities.

Notice that the relevance of the problem of an effective exploitation of web 2.0 tools for learning and knowledge
management purposes appears to be widely recognized (Seuss, 2011). A number of tools is currently available,
mainly for the academic community, allowing users to search for and organize both human resources and
documental resources. Mendeley (http://www.mendeley.com) is one such instance. Their present limit is that they
are not semantic tools; so their search power is limited to formal documental resources and to self-defined user
profiles, which could give rise to ambiguities and inconsistencies.

As it appears in our test case, problems of semantic inconsistency of the research key syntax in different languages
have been recognised in LinkedIn and similar problems might be present in other social networks.

References


Research, 32(1), D267-D270.

Research, 31(6), 533-544.

457.


Palmër, & D. Müller. (Eds.), Proceedings of 2nd Workshop Mash-Up Personal Learning Environments (MUPPLE'09) (pp. 6-14).
Nice, France: CEUR Workshop Proceedings.

Open Learning: The Journal of Open and Distance Learning, 25(2), 141-151.


approaches for mash-up personal learning environments in formal and informal learning settings. In F. Wild, M. Kalz, M. Palmër,
& D. Müller (Eds.), Proceedings of 2nd Workshop Mash-Up Personal Learning Environments (MUPPLE’09) (pp. 23-30). Nice,
France: CEUR Workshop Proceedings.


Proposition and Organization of an Adaptive Learning Domain based on Fusion from the Web

Mohammed Chaoui and Mohamed Tayeb Laskri

Department of Computer Science, Badji Mokhtar University, Annaba, Algeria // chaoui.mohamed@yahoo.fr // laskri@univ-annaba.org

ABSTRACT

The Web allows self-navigated education through interaction with large amounts of Web resources. While enjoying the flexibility of Web tools, authors may suffer from research and filtering Web resources, when they face various resources formats and complex structures. An adaptation of extracted Web resources must be assured by authors, to give reliability, satisfaction of learners and content quality of E-learning platform. This study has proposed Fusion of Web resources approach to this problem and organized resources to an Adaptive Learning Domain into E-learning platform. The focus was firstly, on searching tools and filtering methods to extract the most relevant educational Web resources and structuring them to create courses, secondly, on adaptation of extracted Web resources. Our approach explores a new process of fusion in creation and in adaptation to learner’s profiles. That approach doesn’t need much time and many efforts which can be taken and done by authors to create courses. It also finds direct way to Web resources which are needed, and the update of courses that can be done directly from the Web with the reuse of extracted resources. The evaluation of this work has given high performance in comparison with different methods of course's creation.

Keywords

Web resources, Adaptive Learning Domain, Fusion, E-learning Platform, and Learner Profiles

Introduction

Technology has created cyberspace and decomposed national and international barriers. Today anyone possesses computer, modem and supplier access which can be connected to the new means and enjoy its many gifts. Everyone can teach and anyone can learn what, when and where he wants. Nowadays, the rapid technological revolution has changed the world totally, especially at the level of knowledge and information’s transfer inside and outside the university, thus gives new type of higher education fundamentally different from the previous one (Brusilovsky, 1999). Online education as a tool for distance learning is considered as a teaching method that allows teacher and learner to be separated from each other in time and in place while they are remaining connected by an appropriate technology (Kolb, 1984). In online education, the separation time and space is filled by ICT (Information and Communication Technologies) networks, including the Internet and World Wide Web. E-learning is useful in education, business and all types of learning. Through this method, money and time can be saved. In addition to, producing measurable results (Amaral & Leal, 2006; Arthurs, 2007; Bostrom, Olfman & Sein, 1990; Brusilovsky, 2004). E-learning is mainly the network transfer of skills and knowledge in E-learning platform by using electronic applications and learning process (Brooks, Kettel & Hansen, 2005; Rodic & Anca, 2008).

With the rapid development of digital technology, information has become a very important area of research through which researchers open doors on several researching areas, such as indexing, backup and representation of information (Senan, 2008). And then, it has a responsible way for each course preparation in online training, but it also requires some steps, especially the selection of the most relevant information (Chaoui & Laskri, 2011b). Information is available in several formats on the Web, particularly in text format flexible to any changes and can be updated also by authors and learners. This change and update make the task of adapting the information very difficult and not completed process (Canales et al., 2007; García-Barrios, 2006; Monova-Zheleva, 2005; Rosmalen et al., 2006; Ruiz et al., 2008). Additionally, the amount of learning materials on the Internet has grown rapidly in recent decades. Therefore, the information’s consumers are challenged to choose the right things. In E-learning systems, most approaches have led to confusion for learners. For that, adaptive learning has gained much attention in this area (Chaoui & Laskri, 2011a; Monova-Zheleva, 2005; Papanikolaou et al., 2002; Rey-Lopez et al., 2009; Sampson, Karagiannidis & Cardinali, 2002; Wang, Wang, & Huang, 2008; Yang & Wu, 2009; Yessad et al., 2008).

Our aim through this work PrOALDoF-Web “Proposition and Organization of an Adaptive Learning Domain based on Fusion from Web” to reduce the huge space of the Web which contains billions of Web pages, that huge space becomes personalized one with direct learners’ adaptation to meet their satisfaction and to provide good training which can be changed and updated, but with reliable and academic resources.
The rest of the paper is organized as follows: in the first part; we give background and related works. In the second one we detail our project "PrOALDoF-Web", its algorithm and different steps to create courses, justifying our ideas and the new approaches and developed methods throughout the implementation of PrOALDoF-Web. The importance of our project is the adaptation and is not the definition of learner’s model for tracking learners and subsequently the decision of learners’ levels in order to adapt the appropriate courses, but we assume that learners’ levels are introduced into the system. The third part is evaluation in comparison with our approach and different methods of courses’ creation.

**Adaptation in E-Learning Systems and adaptation in our Approach**

The point which makes E-learning more practical than the traditional methods of learning is: E-learning can be addressed to the maximum number of participants with diversity of learning styles, preferences and needs. For this, any adaptive E-learning system must have high quality of training (Akoumianakis, 2011; Lee, Hsieh & Hsu, 2011). In addition to, E-learning has advantages such as reducing overall costs and time, proving that completion and certification are essential elements of training initiatives such as: interactivity and the ability to deliver consistent contents which are needed according to demands (Kruse, 2004). These E-learning environments have made satisfaction to the demands of learners, who have a central role in learning, for that these environments have become increasingly popular (Chen, 2009; Lee, Hsieh, & Hsu, 2011; Monova-Zheleva, 2005; Rosmalen et al, 2006, Wang et al., 2010). The main point of such popular E-learning system which must be put under consideration is to find out the learner’s preferences, interests and browsing behavior to offer adapted training (Chen, 2009, Wang et al., 2011). This consideration highlights an idea of managing an adaptive e-learning system (Brusilovsky, 1999). This method is considered as an alternative to traditional methods and opens a door for new approaches in the development of educational courseware (Akoumianakis, 2011; Caravantes & Galán, 2011; Peterson, Rayner & Armstrong, 2009; Lee, Hsieh, & Hsu, 2011; Surjono, 2009; Wang et al., 2010; Wang et al., 2011).

Much effort has been made in the field of adaptive systems to offer good user’s model in learning systems (Nguyen & Do, 2008). Most of these works are about the learning methods of learners to gain more details about learner's needs (Liegte & Janicki, 2006; Magoulas, Papanikolaou, & Grigoriadou, 2003; Stach, Cristea & De Bra, 2004; Yang & Wu, 2009). There are two different general approaches to the adaptation of learning content (Olfman, Mandviwalla, 1994; Papanikolaou et al., 2002). The first approach seeks to adapt the learning content with special needs, and the second one focuses on the provision of the most appropriate learning content to the learner’s needs. The first is called the adaptation of the content’s level and the last one is called the link-level adaptation. Neither approach can be preferred from the other one. Based on these approaches, several projects of researches have given a new methodology for appropriate content. Some of these projects are oriented to the extension of the standard learning content to improve the quality of the learning process. One group argues that current standards do not support an adaptive system, so they must be modified in some respects (Lu & Hsieh, 2008; Rey-Lopez et al, 2009; Sampson, Karagiannidis & Cardinali, 2002). In response to the fact that the metadata standards of learning content are somehow inadequate for some applications, a group of researchers have tried to replace them by the ontology that considered one of the pillars of the Semantic Web (Chi, 2009; Jovanovic et al., 2007; Lee, Tsai & Wang, 2008; Shih, Yang & Tseng, 2009; Verbert et al., 2005; Wang & Hsu, 2006; Yang, 2006; Zitko et al. 2009). Researchers have developed systems based on Semantic Web by creating an ontology used for the themes of the course. This method of modeling permits interaction between learners and systems as which have been done in the researches by (Chi, 2009, Jovanovic et al, 2007; Lee, Tsai & Wang, 2008; Shih, Yang, & Tseng, 2009; Verbert et al., 2005; Wang & Hsu, 2006; Yang, 2006; Zitko et al., 2009). There are some studies that have used a multi-agents system in adaptive learning (Canales et al., 2007; Chen, 2008). Four types of agents exist in that proposed system: The first one does the management of context, the second one does the selection of content, the third one does the organization of content and the fourth one does the presentation of content.

Our objective is to develop new approach in adaptive E-Learning systems that can be integrated with any LMS. Learning content can be reused and shared across different platforms. The source of information of that content is the Web. The method of research and filtering is based on ontology and semantic rules. The adaptation approach is based on fusion via ontology of domain and pedagogical ontology. What is new in our work is the fusion of several fragments of Web resources to increase the quality of training content and to get adaptive, reliable, rich and dynamic system; dynamic is in the sense enrichment and update.
Proposed Approach of “PrOALDoF-Web” System

In this part, we have detailed our proposed approach (as in Fig 1.) to construct the different parts of our system, and to give new tool with high performance in research and filtering of Web resources based on Degree of Relevance ‘DR’ and Distance Based on Semantic Rules ‘DBSR’. The two methods are based on ontology of domain. We have organized our course from hierarchy of ontology of domain to construct Learning Domain ‘LD’. After that, we have got Adaptive Learning Domain ‘ALD’ through the application of pedagogical ontology. Thus, contains learner levels and strategy of adaptation based on fusion. ALD presents database that contains the most relevant Web resources and with the application of fusion approach, the course content will get a high quality and the possibility of the reuse of the extracted Web resources.

![Figure 1. Proposed Approach PrOALDoF-Web](image)

Beginning

- Searching resources in the Web via Google API, with an automatic formulating of query based on ontology of domain; we have defined for each concept some keywords.
- Filtering based on domain ontology; we calculate degree of pertinence “DP” in the first round, to reduce space of filtering. In the second round, we determine the distance between sub fragments in each fragment of each document or resource in the Web, the distance is based on semantic rules “DBSR”.
- Selecting the most relevant resources.
- Saving fragments in the New fragments Database "NFDB".
- Fusion of sub fragments content based on distance between sub fragments and semantic rules (Distance Based on Semantic Rules) "DBSR".
- After the fusion we get a new fragment which can be associated to a part of the course.
- Saving new fragments in New fragments Database "NFDB";
- Connecting “NFDB” database with pedagogical ontology to create learning domain.
- Connecting the learning domain with learners.
- Adapting course to learners to create our Adaptive Learning Domain Based on Fusion from the Web.
- Enrichment of “NFDB” database from the Web;
- Resources of the Adaptive Learning Domain can be reused directly from “NFDB” database.

End.

![Figure 2. PrOALDoF-Web Algorithm](image)
To create adaptive course in PrOALDoF-Web System, we have five steps: (1): Research in Web, (2): Filtering and Selection, (3): Fusion of sub fragments, (4): Construction of Learning Domain and (5): Adaptation to Learner's Profiles. To better understanding the proposed approach, we have realized the algorithm called “PrOALDoF-Web” (as in Fig 2.). We have detailed principal of PrOALDoF-Web and what is new in our project shows the difference between our work and similar projects. In the first time, method began with filtering of Web resources by applying statistical and semantic methods based on concepts of ontology and on a list of key-words for each concept, to define list of the most relevant documents for each part of a course. After that, we apply semantic method (according to semantic rules) for each fragment between terms which are key-words of each concept. For this, the two methods are complementary to give list of fragments for each concept that presents part of a course just after extracting the most relevant fragments, in online way. Those fragments are stored in database; that is to say, our two methods have given results without downloading documents, so that to minimize the time.

With the fusion approach, course can be created from different fragments that are extracted from one or more Web resources. Adaptation of the course content is based on fusion. The course content can be changed according to learners’ levels. When the level increased, the enrichment of course increased too.

(1): Research in Web

We have used Google API automatically in our system instead of using manual search through Google search engine. This API gives list of results ‘URLs’ thus can be obtained from queries. These queries are formulated automatically from keywords and annotations which are located ontology of domain, in each concept. Google API permits to filter results by types of documents (PDF, Word, etc. ...) that are given via ‘filetype’. By the application of Google API in the system, the course part chosen by author will generate a query. An example of research can be given: Course is about Human Skeleton, Part = Head => Query = Head in Human Skeleton filetype:doc. Head in human skeleton: are keywords and “filetype:doc” specifies the criterion of filtering documents. In this example, the result is a list of URLs and documents are word (.doc) type. After that, we have used our two methods of filtering ‘DP’ and ‘DBSR’. We give further details in next part.

(2): Filtering and Selection

We have used JENA API which is functioned only in JAVA language. JENA API gives a possibility of integrating the ontology of domain (as in Fig 3.) in the system. With this API, We can read and write ontology as in these types (OWL or RDF …and so on). In our case we have OWL type. The hierarchy of the course is obtained from ontology of domain (as in Tab 1.).

\[Figure 3. \text{A graphical extract of the created ontology of domain}\]
Table 1. Hierarchy of the course from ontology of domain

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>To Concept L</th>
<th>=&gt;</th>
<th>Chapter 1</th>
<th>Chapter 2</th>
<th>To Chapter A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub Concept 1</td>
<td>Sub Concept 2</td>
<td>To Sub Concept M</td>
<td>=&gt;</td>
<td>Subtitle 1</td>
<td>Subtitle 2</td>
<td>To Subtitle B</td>
</tr>
<tr>
<td>L: the last concept in ontology</td>
<td>M: the last Sub concept of the correspondent father concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We have prepared with this hierarchy our Learning Domain to save, in the next time, the extracted fragments in correspondent parts of the course in new fragments database ‘NFDB’. New fragments are due to the process of sub fragments fusion. ‘NFDB’ database has new Excel model (as in Table 2.).

Table 2. Portion of Excel Model to create ‘NFDB’ Database

<table>
<thead>
<tr>
<th>Part 1</th>
<th>SR1</th>
<th>SF1</th>
<th>DR</th>
<th>DBSR</th>
<th>SF2</th>
<th>DR</th>
<th>DBSR</th>
<th>SFN</th>
<th>DR</th>
<th>DBSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSF1</td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td>SRN</td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td>Part 2</td>
<td>SR1</td>
<td>SF1</td>
<td>DR</td>
<td>DBSR</td>
<td>SF2</td>
<td>DR</td>
<td>DBSR</td>
<td>SFN</td>
<td>DR</td>
<td>DBSR</td>
</tr>
<tr>
<td>FSF2</td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td>SRN</td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td>Part N</td>
<td>SR1</td>
<td>SF1</td>
<td>DR</td>
<td>DBSR</td>
<td>SF2</td>
<td>DR</td>
<td>DBSR</td>
<td>SFN</td>
<td>DR</td>
<td>DBSR</td>
</tr>
<tr>
<td>FSF3</td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
<tr>
<td>SRN</td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.0</td>
</tr>
</tbody>
</table>

In addition to use a Java Excel API; this API allows reading and writing an Excel document in Java Platform. For each part of course, we have defined some semantic rules ‘SRs’ to calculate ‘DR’ and ‘DBSR’ of each sub fragment ‘SF’ which can be found in one or multiple Web resource parts. The SRs of each course part are defined and organized vertically in the table below. And the SRs have also sub fragments that are organized horizontally; these later ones are extracted from Web resources.

After this, we have started fusion for each course part in the table. For example Part 1: we have chosen the most relevant contents which are stored in sub fragments from ‘SF1’ to ‘SFN’ for each SR, as a result from fusion we get new fragments that must be saved in the correspondent column which is called ‘FSF’ ‘Fusion of Sub Fragments’.

To obtain results of the Table 2, we must complete series of steps as follows: First, we have divided Web resources found into sections or paragraphs as follows: <P1>, <P2> … <PN>; PN is the last part of Web resources found. Then, we have used fragmentation on each fragment, the aim from that is to split them into sentences and into separate words. And then we have eliminated needless words (the, of, a, an, or, in, etc. …), So that these ones do not affect the result of two methods ‘DR’ and ‘DBSR’ that are determined below. To create course parts, we have to extract the relevant fragments in Web resources found. For that reason, we must calculate ‘DR’ in the first time to determine relevance of the appropriate course part. In the second time, we must calculate ‘DBSR’ only on the most relevant fragments to extract sub fragments that are related to the SRs defined for each course part. We can extract sub fragments from multiple Web resources fragment. Those sub fragments can be given together to one course part or dispatch them to multiple course parts. ‘DR’ and ‘DBSR’ are two original and proposed methods in our work.

Degree of Relevance ‘DR’

It is a statistical method (1), based on the frequency of ontology concept (which presents a component of the course) in Web resource fragment at the first time and the existence of keywords and their frequencies in the same fragment.
Drat the second time. The frequency of word in one fragment is how many times the word can be repeated in this fragment. The formula that is mentioned below used to calculate the ‘DR’ in one Web fragment.

\[
DR = \frac{F_c + \sum_{k=0}^{n} (F_k \cdot W_k)}{N} \quad (1)
\]

Where: c: Concept. \( F_c \): Frequency of concept c in one Web resource fragment. k: Keyword. n: Maximum number of keywords. \( F_k \): Frequency of keyword k in one Web resource fragment. \( W_k \): Weight of keyword k. N: Total number of words in one Web resource fragment.

\[\sum_{k=0}^{n} (F_k \cdot W_k): \text{Sum of frequencies of keywords (k=0…n) for one ontology concept in one Web resource fragment, multiplied by weight of correspondent keyword.}\]

To say that a fragment is relevant to one course part, we must calculate its degree of relevance over all course parts. The fragment which has the highest degree will be saved in ‘NFDB’ database, and then we repeat the process for all fragments of Web resources found.

**Distance Based on Semantic Rules ‘DBSR’**

It is a semantic method (2) based on the distance between terms in sub fragments. Firstly, we must extract terms from one sub fragment and calculate distance only between terms that are defined in semantic rules ‘SRs’. ‘DBSR’ presents a projection of SRs on sub fragments of the Web resource found just to extract the most relevant appropriate sub fragments to one part of the course. When we have successive terms in sub fragment of Web resource, and some distances equal 0, ‘DBSR’ result will be near to 1. As consequence, we have to associate the sub fragment to ‘SR’. In the case of not finding sub fragments to one sub part of the course, we have to use a dictionary of synonyms terms to give more possibilities to get the needed sub fragments that contains different information defined in SR. A formula’s mentioned below for calculating "DBSR" in one sub fragment of Web resource.

\[
DBSR = \frac{\sum_{t=0}^{n} (\text{if } E_t \text{ in SF then } 1 \text{ else } E_{st}^n \text{ in SF})}{N} \quad (2)
\]

Where: t: Term in SR. n: Total number of terms in SR. \( E_t \): Existence of term t in Web resource sub fragment. SF: sub fragment. st: synonym of term t. stn; total number of synonyms terms. \( E_{st}^n \) in SF: Existence of all synonyms of term t in Web resource sub fragment.

\[\sum_{t=0}^{n} (\text{if } E_t \text{ in SF then } 1 \text{ else } E_{st}^n \text{ in SF}): \text{Sum of existences of terms or synonyms in one Web resource sub fragment.}\]

We have repeated the same treatment for all Web resource sub fragments, to associate them to SRs in course part, and saved results in ‘NFDB’ database. After that, we must realize fusion process with sub fragments extracted from Web resources. We will give more details in the next part of paper: Fusion of sub fragments.

We have obtained comprehensive approach that meets our needs:

- The hierarchy of course is taken out from ontology of domain.
- Annotations and keywords of each concept in ontology assure calculation of ‘DR’ (1) of each fragment extracted from Web resources to find the most relevant fragments of course parts. We calculate ‘DBSR’ (2) for all relevant fragments to extract the most relevant sub fragments of course parts.
- Finally, we order the most relevant sub fragments in Excel Model to create our New Fragments Database ‘NFDB’.

**(3): Fusion of sub fragments**

Fusion is based on semantic rules 'SRs' between terms in each sub part of the course (as in Fig 4.) (One or more concepts have many SRs). We have created SRs to all course parts. Each ones has been associated to one part of the course.
Ontology of domain plays three roles: the first one is the presentation of the entire hierarchy of the course, as we have mentioned in the title (Filtering and Selection), it allows calculation of ‘DR’ (1). The second one is the backup of ‘SR’, which allows the calculation of distances ‘DBSR’ (2) between terms in each part of the course. The third one is to offer the possibility of the fusion of the fragments found after research, filtering and selection.

The fusion can take place in the two following cases: the fusion of one or more Web resources and the fusion between sub fragments in the same Web resource. In these two cases, we can find sub fragments in several places in the Web resource. There are two different levels of extraction (as in Fig 4.): (1) fragment level which is based on ‘DP’; we apply the method in the first time on all Web resources to find the most relevant for the course as in the example “D1, D2 & D3 are three documents in Fig 4.” After that, we apply the method ‘DP’ for another time for all most relevant Web resources to extract the most relevant fragments as in the example “F1, F2 & F3 of D1 in Fig 4”. (2) sub fragment level that’s based on ‘DBSR’ which presents distance between words of Web resource sub fragments are compared with terms of SRs. We apply method to all extracted fragments to extract the most relevant sub fragments and then we have just associated the green ones to correspondent SRs as in the example sub fragments 1 and 2 are selected “SF1, SF2 & SF3 of F1 of D1”. As a result, we have extracted three sub fragments: D1F1SF1 / D1F2SF1 / D1F3SF1 of Semantic rule one ‘SR1’ in part one ‘P1’; part one of course contains four SRs, the first one has three most relevant sub fragments, and among these three ones we must choose D1F1SF1 (Sub Fragment one of Fragment one of Document one). After that, we have selected sub fragments with bold color to get components of each part of the course. Results are saved in ‘NFDB’ database. We can illustrate the process of that fusion by using this example: Fusion of three most relevant fragments in three different documents to the same part of the course (as in Fig 5.). In this example, we have selected from ‘NFDB’ database sub fragments which have been associated to SRs. There are three parts of the course which have different colors (Blue, Red and Yellow).
Figure 5. Fusion process of three Web resources

(4): Construction of Learning Domain

After the process of fusion, we get all sub fragments which are saved in ‘NFDB’ database (as in Fig 4.). For this ‘NFDB’ database presents our Learning Domain ‘LD’. ‘LD’ can also be adapted to learner profiles, because our pedagogical ontology is based on semantic rules defined in ontology of domain, that is why each sub fragment extracted from Web resource is associated to one SR which gives us one part of the course. We will mark the process of adaptation in the next part of the paper (Adaptation to Learner’s Profiles).

(5): Adaptation to Learner’s Profiles

At this moment, we have not a learner model but after that we can use its outputs (learners’ level). We have created a method which is based on learners’ test. Through this method, we can determine the level of each learner. In addition to the creation of a pedagogical ontology which may be used when we have a learner’s level. In our approach, we have enriched pedagogical ontology with same semantic rules defined in ontology of domain, but grouping the SRs in levels; that is to say, each level has some numbers of parts of the course, and each part has some numbers of SRs associated with it. When the level of learner increased, the numbers of parts increased too (as in Fig 6.). At the end of our work, we will reach the realization of the Adaptive Learning Domain from Web resources.
We have associated each 'SR' for one level like (SR1 L1: Semantic Rule N°1 is associated to Level N°1). And for each part of the course, we have to associate each SR to only one level as in this example: (P1 “Part of course N°1” contains four (04) 'SRs' “SR1 L1, SR2 L1, SR3 L3 & SR4 L3”). With the fusion of different 'SRs' of one level or multiple one, as consequence, we can get sub fragments of learner’s level. Level number one is composed of three sub fragments “D1F1SF1 (Sub Fragment one of Fragment one of Document one), D1F2SF2 (Sub Fragment two of Fragment two of Document one) & D2F1SF3 (Sub Fragment three of Fragment one of Document two)”. The extraction of sub fragments is based on fusion process that is explained in the precedent part of paper.

**Evaluation of PrOALDoF-Web System**

We have two points: Items of comparison and Results of Comparison.

(1) **Items of comparison**

We have 14 points or items to evaluate the quality and usefulness of our approach. To get more benefit, we have compared PrOALDoF-Web with three ways of courses’ creation: manual creation by authors, creation of courses in E-Learning platforms by authors and the tools of creation. The last one can be divided into two groups: Commercialized Platforms (MOODLE …Etc.) and recent researches in E-Learning Platforms. We have obtained 9 items from the Learning Object Review Instrument (LORI 1.5) that is an evaluation framework for multimedia
learning resources (Nesbit, Belfer & Leacock, 2004) as it’s explained in Tab 3. Beside that, five items are proposed in process of evaluation of PrOALDoF-Web, as in Tab 4.

<table>
<thead>
<tr>
<th>Table 3. Items in LORI 1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td>Content quality</td>
</tr>
<tr>
<td>Learning goal alignment</td>
</tr>
<tr>
<td>Feedback and adaptation</td>
</tr>
<tr>
<td>Motivation</td>
</tr>
<tr>
<td>Presentation design</td>
</tr>
<tr>
<td>Interaction usability</td>
</tr>
<tr>
<td>Accessibility</td>
</tr>
<tr>
<td>Reusability</td>
</tr>
<tr>
<td>Standards compliance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. Items proposed in PrOALDoF-Web</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td>Method of creation</td>
</tr>
<tr>
<td>Enrichment method</td>
</tr>
<tr>
<td>Ease of Use</td>
</tr>
<tr>
<td>Content Research</td>
</tr>
<tr>
<td>Experience of Authors</td>
</tr>
</tbody>
</table>

In our study, we have three methods of courses' creation "Manual, Semi-Automatic and Automatic method"

(1): **Manual Method**: used directly by the authors, they seek information in a domain, using different methods and engine of research to choose the most relevant information in different resources of it (Internet, books, papers ...Etc). The authors must filter results in order to create their own course in electronic format (Word, PDF or another format).

(2): **Semi-Automatic Method**: There are two sub categories: Assisted Method and Dynamic Method

- **Assisted Method**: a method which can be assisted by an author, using tools of courses’ creation such as: LMS "Learning Management System" or LCMS "Learning Content Management System" that are available in the training platforms as: (Moodle, Sakai, dotLRN, OLAT, Olat LMS, OLMS, …etc). There are also other ones such as: free software (Adobe Coursebuilder, CALI Author, CourseLab, eXe, AuthorLCDS, Learning Content Development System, Lesson Writer, Memorize, MOS Solo …etc.) or commercial software (Adobe Captivate, Articulate, Camtasia, Composica, Outstart Trainer, OpenWorld Presenter, PowerTrainer Authoring Tool …etc.) which are available in the market. In these cases mentioned above, authors must search for information and filter results and extract the most relevant information to fill all fields in process’ creation.

- **Dynamic method**: following the pedagogical aspect to formulate bricks of information that are previously created by authors through proposed model to achieve the goal which is the adaptation of courses to learners’ profiles. Here, there are examples which can illustrate the recent research works, among them there are those proposed approaches based on: ontology, multi-agent systems ... etc.
(3): Automatic Method: Through PrOALDoF-Web system that offers an approach which eliminates the lack in the first two cases (Manual and Semi-Automatic method). This lack can be found in the search and the filtering of information when the course is being created. Moreover the approach provides an opportunity to increase the quality of courses by combining different resources, and then gives the advantage of adaptation of content to learners’ profiles according to the proposed approach for the desired target.

(2): Results of comparison

We evaluate the proposed approach. Firstly, we have validated ontology of domain with three specialists. Secondly, series of experiments and analyses were performed. At the initial stage of the project, the evaluation was on author perceptions and reactions towards the system and its proposed functions of fusion of Web resources and the adaptation to learners’ profiles. After that, the proposed approach has been properly implemented to the efficiency of that system. Then the evaluation was on learners’ perceptions and surveys about system. 127 students and 17 teachers from University of Annaba-Algeria have participated in this evaluation by using the JAVA PrOALDoF-Web system. The results of students are as follows: 84.25% of satisfaction, 12.59% of dissatisfaction and 3.14% which are other views. In addition to, the results of teachers are as follows: 76.47% of satisfaction, 11.76% of dissatisfaction and 11.76% which are other views. The evaluation is related to the levels of comparison that are detailed in Table 5.

Table 5. Levels of Comparison

<table>
<thead>
<tr>
<th>Points of Comparison</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method of creation, Enrichment method, Content Research</td>
<td>Manual</td>
<td>Under Medium</td>
<td>Medium</td>
<td>Over Medium</td>
<td>Automatic</td>
</tr>
<tr>
<td>Ease of Use, Experience of Authors, Content quality, Feedback and adaptation, Learning goal alignment, Motivation, Presentation design, Interaction usability, Accessibility, Reusability, Standards Compliance</td>
<td>Low</td>
<td>Under Medium</td>
<td>Medium</td>
<td>Over Medium</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 6 provides a comparison with PrOALDoF-Web in automatic way and with different methods of courses’ creation by authors in manual way (C1, C2 & C3 in table 6 present three courses of three different authors) or with authoring tools in semi-automatic way. Number one to five ‘1...5’ of each column in Table 6 is explained in Table 5.

Table 6. Results of Comparison

<table>
<thead>
<tr>
<th>Points of Comparison</th>
<th>Tool of Creation</th>
<th>Our Approach PrOALDoF-Web (Automatic way)</th>
<th>Manual creation (By Authors)</th>
<th>Creation of Course in E-Learning platforms (By Authors with tools’ creation)</th>
<th>Commercialized Platforms (MOODLE...Etc.)</th>
<th>Research Platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method of creation</td>
<td>5</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Enrichment method</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Content Research</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Experience of Authors</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Content quality</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Learning goal alignment</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Feedback and adaptation</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Motivation</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Presentation design</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Interaction usability</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Accessibility</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Reusability</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Standards compliance</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Method of creation, Enrichment method and Content research are three different points of comparison. These three points work automatically in PrOALDoF-Web. The first interest of our approach is to create courses directly from Web resources and to save time. The strategy has allowed to add other sub fragments in database, for the reason that
our main goal is to find for each part of the course a list of pertinent documents from the Web, and the content of course will be created from the extracted sub fragments which have a perfect relevance. However, in the other methods, the three different points of comparison which are mentioned above work either manual or semi-automatic. These previous ways of creation can take much time when it’s purely manual, in contrast to using tools’ creation through which time can be saved.

**Ease of use, Experience of authors and Content quality:** In our system and in the most last researches about tools’ creation, the use of these different components is easier, each part of the course can be controlled and validated by experts of domain who have enough experiences, as consequence courses may have higher quality.

**Feedback and adaptation / Learning goal alignment / Motivation:** are three different points of comparison. In our system, the adaptation is assured by the approach which is based on fusion, because we have the goals which are defined in pedagogical ontology for each learner and the content that can be used in creation of courses is stored in ‘NFDB’ database. In the most last researches about tools’ creation, experts have proposed approaches which can do the adaptation which cannot be done by the other tools’ creation. In the first cases above, the systems have a specific ability to motivate and interest learners.

**Presentation design and Accessibility:** The two points of comparison have not been completely assured by our system; they can also perform a future enrichment in this field. There are many works in this area where researchers have given the quality of presentation design and satisfied accessibility in platforms for learners. Those two points of comparison cannot be done manually.

**Interaction usability, Reusability and Standards compliance:** are three different points of comparison. In our system and in the most last researches about tools’ creation; the navigation in system is easier, the quality of interface is higher. In addition to a help that can be used by learners. Systems can be used in varying learning contexts for different learners. These two precedent cases of creation assure the international technical standards and specifications. Those three points of comparison cannot be done in the other tools’ creation.

To better understanding the results of comparison, we have made the diagrams of the six cases about courses' creation with the 14 points of comparison (as in Fig. 7.).

![Figure 7. Diagrams of the different methods of courses’ creation](image-url)
Discussion and Conclusion

The results of evaluation reflect the success of PrOALDoF-Web system talking about the performance of items in comparison with our system and different methods of courses’ creation (methods which have been mentioned in our work). Through the study which was carried out with: (Research, Filtering and Selection) of Web resources, Fusion of sub fragments and Adaptation. All those steps that are cited, they have been used in order to construct an Adaptive Learning Domain that can be adapted according to learners’ profiles. This study gives us a complete system of courses’ creation and the adaptation of courses which are designed for learners. The proposed approach has offered new techniques which have made the Web as a medium of resources which are well organized and adapted directly to learners’ needs. Adaptation of extracted Web resources gives reliability, satisfaction of learners and content quality in E-learning platform. This approach doesn’t need many efforts and much time which can be done and taken by authors to create courses which can get a high quality because their contents are obtained from the large richness of Web. The update of courses is done directly from the Web with the reusability of the extracted resources.

Through this presented research, we have achieved the idea that the different methods of courses’ creation which cannot be assured automatically, because those methods offer either interfaces where authors can download their courses or tools which give texts zone that must be filled by authors who are implicated in such creation and they have their own scope of research about resources in the Web (search engines, online libraries, sites of universities etc.). Therefore, authors should provide great effort to filter resources and extract the contents which are needed for their courses. PrOALDoF-Web is completely based on Web for the objective to create courses in automatic way using ontology of domain and semantic rules which may present some difficulties in their manual creation, because for each fragment of course we create an important number of SRs. This step needs more efforts by experts. The efficiency of the approach will be improved when the number of SRs increased. For that reason we have to look for an automatic creation of SRs.

References


Revisit the Effect of Teaching and Learning with Technology

Yuan-Hsuan Lee¹, Hersh Waxman², Jiun-Yu Wu¹, Georgette Michko³ and Grace Lin⁴

¹Institute of Education, National Chiao Tung University, Hsinchu City 300, Taiwan // ²361 Harrington Office Building, Texas A&M University, College Station, TX, USA // ³111 CHC, 4800 Calhoun Road, the University of Houston, Houston, TX, USA // ⁴1776 University Avenue Wist Hall, University of Hawaii, Honolulu, HI, USA // jasvi.rms@gmail.com // jiunyuwu@mail.nctu.edu.tw // hwaxman@tamu.edu // gmichko@uh.edu // gracelin@hawaii.edu

(Submitted May 3, 2011; Revised August 4, 2011; Accepted October 28, 2011)

ABSTRACT

We re-examined the effect of teaching and learning with technology on student cognitive and affective outcomes using the meta-analytic technique. Screening studies obtained from an electric search of databases such as PsyInfo and ERIC resulted in 58 studies (1997-2011). Overall, effect sizes were small to moderate across the cognitive and affective outcome measures. Specific teaching/learning components such as context/making sense, challenging activity, instructional conversation, and joint productivity were associated with effect sizes. Instructional features such as objectives, pattern of student computer use, and type of learning task also moderated effect sizes. Suggestions are made for pre-service teachers and in-service teachers to include these instructional features and teaching strategies in teaching and learning with technology.

Keywords
Technology, Teaching and learning, Computer-assisted instruction, Good instructional practices, Student outcomes

Introduction

Integrating technology into classroom teaching and learning has been an important issue in the last few decades. Several meta-analyses have been conducted to examine specific modes of instruction or educational practices that promote student learning and teaching with technology. Lou, Abrahim, and d’Apollonia (2001), for example, examined the effects of small group versus individual instruction with technology and found that small-group learning had more positive effects than individual learning. Other meta-analyses in technology have examined topics such as the effectiveness of interactive distance education (Cavanaugh, 2001), the effect of computer-assisted instruction (CAI) on beginning readers (Blok, Oostdam, Otter, and Overmaat, 2002,), CAI in science education (Bayraktar, 2002,), and the effect of technology on reading performance in grades 6-8 (Moran, Ferdig, Pearson, Wardrop, & Blomney, 2008). A recent meta-analysis by Li and Ma (2010) investigated the influence of computer technology on mathematics achievement in K-12 classrooms from 46 studies and found a greater effect for elementary over secondary school students and that the technology effect was greater when constructivist approach was incorporated in the teaching and learning process (Li & Ma, 2010). A more comprehensive meta-analysis for the effect of technology on learning was conducted using a second-order meta-analytic technique involving 25 meta-analyses encompassing 1055 studies in a 40 year span (Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011). This study yielded a moderate effect size of .35

However, these recent reviews focused only on a particular issue (e.g. group size, CAI, or the general technology effect) and there is little information on what the effective strategies or appropriate approaches are in integrating and using technology in schools and classrooms. For example, Moran et al’s study (2008) found very little research reported outcomes on strategy use and metacognition. Ma & Li’s study (2010), on the other hand, reported a differential effect on constructivist approach versus traditional approach but no specific teaching strategies or instructional features were informed. Likewise, Tamim et al’s study (2011), though very thorough and comprehensive, only included 2 moderator variables on grade level and purpose of technology use. As Ross, Morrison, and Lowther (2010) commented that “educational technology is not a homogeneous ‘intervention’ but a broad variety of modalities, tools, and strategies for learning. Its effectiveness, therefore, depends on how well it helps teachers and students achieve the desired instructional goals” (p. 19), in line with this statement, we aimed to explore what the effective practices are so that teachers and students can teach and learn effectively with technology.
Purpose of this study

The purpose of this study was to evaluate the effects of teaching and learning with technology on student outcomes in K-12 settings so as to inform instructional practices, by reviewing the experimental and quasi-experimental studies published between 1997 and 2011. Unlike prior syntheses, which may focus on a particular teaching practice, grade level, or subject area, we are interested in the overall effect and common technology characteristics, teaching strategies, and instructional features that benefit student learning and teaching across grade levels.

Specifically the meta-analysis intends to address the following research questions:

- What is the general magnitude and direction of the relationship between teaching and learning with technology and student outcomes?
- Are there specific technology characteristics, teaching strategies, and instructional features that affect teaching and learning with technology on student outcomes?

In the following section, we provided a brief review and rationale for the coding of variables based on technology characteristics, teaching strategies, and instructional features.

Technology characteristics

Role of technology and pattern of computer use

Technology can take on several roles in education, such as role of resources, role of delivery system, or productivity. Computer programs were found to be most effective in supporting student centered learning if the programs can provide scaffolds for students with special needs, support factual knowledge acquisition, and emphasize the capacity of technology in creating new learning experiences for students (Pedersen & Liu, 2003). Besides, significant learning gains were found if computers serve as resources (Wegerif, 2004). Pattern of computer use concerns the size of participants working together with technology. Working with technology individually offers greater flexibility for participants to adjust their own pace; on the contrary, working in a larger group (e.g. 6-8 or more) may result in the dominant use of the technology by one or a few persons. Research has shown that students working in small groups (e.g. 3-5) with computers performed better than individual student working with computers (Lou, Abrami, and d’Apollonia, 2001).

Type of technology, software, and objective of technology

Type of technology refers to the carriers (e.g. laptops, PCs, PDAs…etc.) of the instructional material while software is the type of instructional material itself (e.g. tutorial, drill & practice, exploratory environment…etc). For example, laptop programs were found to be effective in student engagement (Penuel, 2006) and academic achievement (Zucker & Hug, 2008). Software, on the other hand, can be very useful if used for an appropriate learning purpose. For example, multimedia talking books can help beginning readers learn to read (Chera & Wood, 2003; Doty, Popplewell, Byers, 2001) and computer simulations can help learn dissection before the actual laboratory anatomy in a biology class (Akpan & Andre, 2000). As for objectives of technology use, technology was found to have a greater effect in learning when used to support instruction rather than for direct instruction (Tamim et al, 2011).

Effective Teaching Strategies

We included teaching strategies as moderators in the meta-analysis. The Center for Research on Education, Diversity, and Excellence developed five standards of effective teaching strategies, namely (1) Teachers and Students Producing Together (Joint Productive Activity), (2) Developing Language and Literacy Across the Curriculum (Language Development), (3) Making Meaning: Connecting School to Students’ Lives (Contextualization), (4) Teaching Complex Thinking (Challenging Activities), and (5) Teaching Through Conversation (Instructional Conversation) (see Dalton, 1998; Tharp, 1997). These standards are based on the best theoretical and empirical knowledge in the field, and there is ample evidence that their use in classrooms may lead to dramatic improvements for the education of all students (Tharp, Estrada, Dalton, & Yamauchi, 2000).
Instructional features

Mode of instruction and role of teacher

Mode of instruction can be discussed in a variety of settings, such as whole-class, small-group, and individualized instruction. Waxman and Huang (1996) found whole-class approaches were frequently observed in lower technology use classrooms where students generally listened to and watched the teacher, while more independent work was observed in classrooms where technology was moderately used. Studies also showed that teachers’ role as facilitator for student learning had a higher effect than as disseminator of knowledge or modeling processes (Dekker and Elshout-Mohr, 2004; Stonewater, 2005; McCrone, 2005).

Task difficulty, type of task, and learning responsibility

Task difficulty is similar to challenging activity to teach complex thinking in the teaching strategy section. Type of task can be for learning basic skills/factual learning, problem solving, project-based learning, or Inquiry/investigation. Project-based learning, for example, was found to have dramatic gains in student academic achievement across states in the U.S. (Thomas, 2000). Learning responsibility can be categorized into teacher-directed, student-centered, system-directed or mixed. Nowadays, there is a trend to call for student-centered learning (Jonassen, 2000).

Methods

For this meta-analytic review, we used selection criteria and review methods that are similar to other recent major national reviews conducted in areas such as teacher preparation (Wilson, Floden, & Ferrini-Mundy, 2001) and reading (International Reading Panel, 2002). The synthesis included quantitative, experimental, and quasi-experimental research and evaluation studies that have been published in refereed journals during a fifteen-year period (1997-2011).

Selection of articles

To be included in the synthesis, articles must satisfy the following criteria:

- Focus on teaching and learning with technology in K–12 classroom contexts where students and their teachers interact primarily face-to-face (> 50 percent of the time);
- Compare a technology group to a nontechnology comparison group, or compare the group at the beginning of the intervention (pretest) to a posttest measure; and
- Have reported statistical data (e.g., t tests or F tests) that allowed the calculation of effect sizes.

First, online computer databases like Education Resources Information Center (ERIC) or PsycInfo were used for search of articles. Keywords such as “Technology/computer” and “achievement” or “technology/computer” and “student outcome” or “technology/computer” and “attitude” were entered in the databases. Over 500 articles were left and met the desired publication time for coverage. Abstracts about these articles were then read to determine if they are relevant to the synthesis. Most of the studies were discarded because they are not comparing the experiment group to a control group that has no access to technology. Other studies were excluded because they are not directly linked with the use of technology for learning and teaching purpose in the K-12 setting. The search and selection procedures resulted in a collection of 58 studies.

Coding design

Our studies were coded on 17 variables. The study descriptors included 2 variables: grade level and publication features (technology journal or educational journal). Technology characteristic descriptors consisted of 5 variables: type of technology, type of software, role of technology, pattern of student computer use, and objective of technology use. The instructional descriptors included 5 variables: learning responsibility, task difficulty, type of
learning task, mode of instruction, and role of teacher. As for teaching strategy descriptors, we included the Five Standards for Effective Pedagogy developed by the Center for Research on Education, Diversity, and Excellence (2002; see Dalton, 1998; Tharp, 1997).

**Interrate reliability**

Two investigators independently coded the studies based on the coding book of 17 characteristics for each of the 366 effect sizes from the 58 studies. Then, each investigator independently coded six studies from the other investigator. The intercoder agreement for each study reviewed exceeded the 85-percent criterion and the average Cohen’s Kappa coefficient reached 0.88.

**Data analysis**

The overall data analysis strategies were based on Lipsey and Wilson (2001). In the initial coding of studies, two types of student outcomes were identified: (a) cognitive, and (2) affective. Effect sizes of standardized mean difference were computed if means, standard deviations, and group size were reported in the selected studies. Otherwise, effects were computed from t-statistics or F-statistics if these were reported. Hedges and Olkin estimator in Lipsey and Wilson (2001) were used to produce unbiased effect size estimates (i.e., Hedge’s g), which are weighted with inverse variance weight (i.e. the inverse of squared standard error value,) so that effects with larger standard error are given a smaller weight because large standard error produces less precise effect size values. In order to insure the independence of ESs, a single combined ES was extracted from each study for each of the two outcomes as suggested by Lipsey and Wilson (2001).

Q-statistics were computed for each outcome based on the adjusted mean effect size weighted with the inverse variance weight function within each study to examine the heterogeneity of effect size (Lipsey & Wilson, 2001). The Q-statistic follows a chi-square distribution with degrees of freedom equal $k-1$, where $k$ is the number of effect sizes (Hedges & Olkin, 1985). Moderators were evaluated using the meta-analytic analog to analysis of variance (Lipsey & Wilson, 2001). In interpreting the Q statistic, a significant $Q_b$ (Q statistic between) suggest a significant mean difference between/among levels of categorical variable, while a significant $Q_w$ (Q statistic within) evaluates the heterogeneity within groups and indicates that a moderator may be needed to group studies into homogenous subcategories (Lipsey & Wilson, 2001). If analysis of moderator effect is needed to investigate sources of variation in effect sizes, a Bonferroni correction with alpha level of .005 will be selected in the analysis to avoid inflated experiment-wise Type I error rate when numerous analyses were conducted for each outcome.

**Results**

A total of 366 effect sizes were computed from 58 studies included in the study. Mean effect sizes were calculated for each construct across studies. For the cognitive outcome, the weighted mean effect (Hedge’s g) was 0.42 with 243 effects from 48 studies. For the affective outcome, the effect was 0.18 with 92 effects from 21 studies.

The chi-square Q statistic was computed for each outcome to evaluate the homogeneity of the mean effects. For the cognitive outcome, $Q (df =47) = 231.47, p< .001$; for the affective outcome, $Q (df =20) = 118.60, p< .001$. The large Q statistics and small p values revealed that the effect sizes were heterogeneous within each construct. Therefore, we conducted analyses of the moderator effect for both outcome measures.

**Analysis of moderator**

*Cognitive*

Results for the cognitive outcome were presented in Table 1. Grade, context/sense making, objective, pattern of student computer use, and type of learning task were significant moderators for effect sizes. Grade 9-12 had the lowest mean effect (.22) compared with grade K-3 (.50), 4-6 (.41), and 7-8 (.59). The finding was similar with Li and
Ma’s study (2010) where secondary schools had a lower effect. Studies showing evidence of making sense or teaching and learning in context (.53) had higher mean effects than those without evidence (.39). Using technology for remediation of skill not learned (.83), finding out about ideas and information (.61), and expressing themselves in writing (.59) had higher effects than for analyzing information (.39), multiple objectives (.19), or others (.26). Studies reporting factual learning (.64), inquiry/investigation (.61), project-based learning (1.39), and others (.62) had higher mean effects than those reporting problem-solving (.39) and mixed type learning (.05). A differential effect was also found on pattern of student computer use. Three to five students per computer had the highest mean effect (1.08) followed by two students per computer (.65), mixed pattern (.57), and others (.44). One student per computer had the lowest mean effect size (.40).

Table 1. Categorical moderators for cognitive outcomes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Qb</th>
<th>dfb</th>
<th>Prob(Qb)</th>
<th>Qw</th>
<th>dfw</th>
<th>Prob(Qw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.42</td>
<td>15.49</td>
<td>3</td>
<td>0.0014</td>
<td>191.76</td>
<td>43</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>GRADE 24=K-3</td>
<td>0.50</td>
<td>25=4-6</td>
<td>0.41</td>
<td>26=7-8</td>
<td>0.59</td>
<td>27=9-12</td>
<td>0.22</td>
</tr>
<tr>
<td>Publication feature 1=technology</td>
<td>7.86</td>
<td>4</td>
<td>0.2408</td>
<td>124.85</td>
<td>27</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Type of Technology 1=PCs .56</td>
<td>2= Laptops .88</td>
<td>3= Networked computer .39</td>
<td>4=Multimedia .61</td>
<td>6=Other .44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software 1=Tutorial .81</td>
<td>2=Drill-and Practice .42</td>
<td>3= Exploratory Environment .38</td>
<td>4=Tools for other task .59</td>
<td>6=Other .41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role of Technology 1=Productivity .41</td>
<td>2=Delivery system .43</td>
<td>3=Resources .45</td>
<td>4=Other .24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern of Computer Use 3=1 student per computer .40</td>
<td>4=2 students per computer .65</td>
<td>5= 3-5 students per computer 1.08</td>
<td>7=Mixed pattern .57</td>
<td>8=Other .44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective 1=Remediation .83</td>
<td>2=Expressing themselves in writing .59</td>
<td>4=Finding out about information .61</td>
<td>5=Analyzing information .39</td>
<td>10=Multiple objectives .19</td>
<td>11=Other .26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context/Making Sense 1=No evidence .39</td>
<td>2=Some evidence .53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenging activities .747</td>
<td>2 .0239</td>
<td>197.77</td>
<td>43</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=No evidence .28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

137
Affective

Results for affective outcome were presented in Tables 2. Challenging activities, instructional conversation, and joint productivity/collaboration were significant moderators for effect sizes. Studies reporting some evidence (.36) or extensive evidence (.25) of challenging activities had higher mean effects than those with no evidence of challenging activities (.06). Likewise, studies with some evidence of instructional conversation (.44) exhibited greater effect sizes than those without evidence of instructional conversation (.12). Those studies showing some evidence (.34) or extensive evidence (.32) of joint productivity/collaboration also had higher mean effects in the affective outcome than those with no evidence (.06) of joint productivity/collaboration.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Q_B</th>
<th>df_B</th>
<th>Prob(Q_B)</th>
<th>Q_W</th>
<th>df_W</th>
<th>Prob(Q_W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>DF</td>
<td>M</td>
<td>SE</td>
<td>V</td>
<td>Sig</td>
<td>95% CI</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>----</td>
<td>---</td>
<td>----</td>
<td>---</td>
<td>-------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>GRADE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24=K-3</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25=4-6</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26=7-8</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27=9-12</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication feature</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=technology</td>
<td>.5645</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=educational</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Technology</td>
<td>10.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=PCs</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Networked-computer</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5=Multimedia</td>
<td>.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6=Other</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role of Technology</td>
<td>7.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Productivity</td>
<td>.0589</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Delivery system</td>
<td>.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Resources</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=Other</td>
<td>.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td>8.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Remediation</td>
<td>.1304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=Finding out about ideas and</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>information</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5=Analyzing information</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7=Improving Computer Skills</td>
<td>1.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10=Multiple objectives</td>
<td>.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern of Student Computer Use</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=One student</td>
<td>.7881</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=Two student</td>
<td>.72091</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7=Mixed pattern</td>
<td>.3019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context/Making Sense</td>
<td>5.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=No evidence</td>
<td>.0201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Some evidence</td>
<td>.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenging activities</td>
<td>12.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=No evidence</td>
<td>.0018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Some evidence</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Extensive evidence</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional conversation</td>
<td>8.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=No evidence</td>
<td>.0033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Some evidence</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint Productivity /Collaboration</td>
<td>14.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=No evidence</td>
<td>.0007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Some evidence</td>
<td>.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Extensive evidence</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Literacy Development</td>
<td>2.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=No evidence</td>
<td>.0871</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Some evidence</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning responsibility</td>
<td>3.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Student-controlled</td>
<td>.2857</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=Teacher-directed</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode of Instruction</td>
<td>14.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=Whole-group</td>
<td>.0056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2=paired</td>
<td>1.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3=Small group (3-5)</td>
<td>.3 .6168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4=Individualized</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

139
Discussion & Conclusions

The main purpose of this meta-analysis was to bring together 15 years of investigations on the effect of teaching and learning with technology on student cognitive and affective outcomes. In terms of magnitude and direction, the overall effect sizes for the two outcomes exhibited a positive effect in teaching and learning with technology. Cognitive outcome, in particular, had an effect size (.42) that was larger than several of the past meta-analytic reviews that were old or covering multiple decades of studies (e.g. Bayrakter, 2002; Christmann & Badgett, 2003; Kulik and Kulik, 1991; Ouyang, 1990; Tamim et al., 2011) but was comparable with meta-analyses analyzing more recent studies (e.g. Li & Ma, 2010; Moran et al., 2008). It is very likely that effect sizes increased with the evolution of technology itself and the advancement of pedagogy in teaching and learning with technology.

Suggestions for pre-service and in-service teachers

Based on the meta-analytic review, we gained invaluable information as to the best practices in teaching and learning with technology. For the cognitive outcome, we found technology was best use for the purpose of basic skills and factual learning which refers to “rote learning and the extent to which participants were able to repeat facts presented during the lesson” (p. 800, Jang, 2008). Factual or rote learning is relatively less complex and less difficult compared to other purposes because it utilizes a more straightforward strategy to learning, such as memorization (Vansteenkiste, Simons, Lens, Soenens, & Matos, 2005). Nevertheless, acquiring basic skills or factual learning is an essential step for students to use technology for other purposes such as expressing themselves in writing, finding out information, analyzing information, and multiple objectives. Our argument can be verified by the fact that project-based learning also yielded the highest effect in terms of type of learning task. The scope of project-based learning usually span across subjects and help learners to see the interconnectedness of multiple domains; it encourages students to search for information, find out about facts, exchange findings, and collaborate with their peers (Kwok & Tan, 2004). Each of these significant knowledge building steps were anchored upon basic skills/ factual learning and instructional elements that are sense-making and contextualized (Arnsen and Saljot; 2007). Therefore, for teachers to improve student cognitive outcomes, the take-home messages are to

- Collaborate in small or paired groups with computers;
- Develop instructional elements that are sense-making in context
- Build student basic skills and help them see the interconnectedness of subject knowledge in a project-based learning

For the affective outcome, collaboration is also an important factor. By working collaboratively, students not only share their cognitive capacity, reduce their mental efforts, but also increase their confidence in the task, which in turn lead to better affective outcome, especially in processing complex tasks (Kirschner, Paas, Kirschner, 2011). The evidence of challenging activity or in some sense similar to task difficulty also contributed to higher student effects. According to the flow theory, people gain their optimal experience in learning and performing when their perceived challenge of task and skill reach a balanced state (Csikszentmihályi, 1990; Moneta & Csikszentmihályi, 1996). In addition, evidence of instructional conversation also promoted student outcome because it is interesting, engaging, focusing on concepts relevant to students, and not dominating by any one student so that extended discussions are found among teacher and students (Goldenberg, 1991). Therefore, the take-home messages for the affective outcome are to include challenging activities, instructional conversation, and joint productivity or collaboration in teaching and learning with computers.

Based on the findings of the study, we would like urge that professional development and teacher preparation be set up with a wide variety of training scopes to include these investigated technology and pedagogical practices for pre-service and in-service teachers in teaching and learning with technology.
References
(Asterisks indicate studies included in the meta-analysis.)


Schacter, J. (2001). The impact of education technology on student achievement: What the most current research has to say. Santa Monica, CA: Milken Exchange on Education Technology.


Research on Education, Diversity, and Excellence.


### Appendix A: Coding information extracted from reviewed studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Grade level</th>
<th>Technology tool</th>
<th>Type of task</th>
<th>Outcome /subject measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akpan &amp; Andre (2000)</td>
<td>7</td>
<td>PCs-simulation</td>
<td>PS</td>
<td>life science</td>
</tr>
<tr>
<td>Alfassi (2000)</td>
<td></td>
<td>NCs</td>
<td>I/I</td>
<td>writing and reading</td>
</tr>
<tr>
<td>Alspaugh (1999)</td>
<td>10-12</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Academic achievement (all areas)</td>
</tr>
<tr>
<td>Bain et al. (2000)</td>
<td>6-8</td>
<td>NCs-hypertext discussion tool</td>
<td>I/I</td>
<td>Literature</td>
</tr>
<tr>
<td>Barker &amp; Ansorge (2007)</td>
<td>4-6</td>
<td>Robotics</td>
<td>PS</td>
<td>Science and technology</td>
</tr>
<tr>
<td>Biggs et al. (2008)</td>
<td>6-8</td>
<td>NCs-interactive software</td>
<td>BS/FL</td>
<td>reading</td>
</tr>
<tr>
<td>Brown et al. (2003)</td>
<td>9-12</td>
<td>NCs - simulation</td>
<td>PBL</td>
<td>academic and technology self-efficacy</td>
</tr>
<tr>
<td>Butzin (2001)</td>
<td>K-5</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Reading, writing and math</td>
</tr>
<tr>
<td>Cady &amp; Terrell, (2007)</td>
<td>5</td>
<td>PCs</td>
<td>BS/FL</td>
<td>self-efficacy attitudes</td>
</tr>
<tr>
<td>Chera &amp; Wood (2003)</td>
<td>K</td>
<td>PCs - multimedia</td>
<td>BS/FL</td>
<td>Reading</td>
</tr>
<tr>
<td>Church &amp; Fisher (2001)</td>
<td>7</td>
<td>NCs</td>
<td></td>
<td>Science</td>
</tr>
<tr>
<td>Cohen (2001)</td>
<td>9</td>
<td>PCs</td>
<td>PBL</td>
<td>Learning style</td>
</tr>
<tr>
<td>Dixon (1997)</td>
<td>8</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Geometry</td>
</tr>
<tr>
<td>Doty et al. (2001)</td>
<td>K</td>
<td>PCs-Interactive story book</td>
<td>BS/FL</td>
<td>reading</td>
</tr>
<tr>
<td>Dybdahl et al. (1997)</td>
<td>5</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Writing</td>
</tr>
<tr>
<td>Erdner et al. (1998)</td>
<td>1</td>
<td>PCs</td>
<td>BS/FL</td>
<td>reading</td>
</tr>
<tr>
<td>Ekane &amp; Maiken (1997)</td>
<td>7-8</td>
<td>PCs</td>
<td>BS/FL</td>
<td>English vocabulary</td>
</tr>
<tr>
<td>Erdogan (2009)</td>
<td>8</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Computer attitude / anxiety achievement</td>
</tr>
<tr>
<td>Estep et al. (2000)</td>
<td>3</td>
<td>Integrated Learning System (ILS)</td>
<td>BS/FL</td>
<td>Math performance, attitude achievement</td>
</tr>
<tr>
<td>Funkhouser (2003)</td>
<td>10-11</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Writing</td>
</tr>
<tr>
<td>Harwell et al. (2001)</td>
<td>6</td>
<td>Tech-integration</td>
<td></td>
<td>Math and computer self-efficacy</td>
</tr>
<tr>
<td>Hertz-Lazarowitz &amp; Bar-Natan (2002)</td>
<td>5-6</td>
<td>NCs-CMC</td>
<td>BS/FL</td>
<td>Higher order thinking</td>
</tr>
<tr>
<td>Hopson et al. (2001-2002)</td>
<td>5-6</td>
<td>NCs</td>
<td>PS</td>
<td>Math and computer self-eficacy</td>
</tr>
<tr>
<td>Isiksal &amp; Askar (2005)</td>
<td>7</td>
<td>PCs</td>
<td>PS</td>
<td>English language</td>
</tr>
<tr>
<td>Keogh et al. (2000)</td>
<td>7-8</td>
<td>PCs</td>
<td>PBL</td>
<td>Cognitive skills</td>
</tr>
<tr>
<td>Ko (2002)</td>
<td>3-5</td>
<td>PCs - computer games</td>
<td>PS</td>
<td>reading, motivation, metacognitive awareness</td>
</tr>
<tr>
<td>Kramarski &amp; Feldman (2000)</td>
<td>8</td>
<td>NCs</td>
<td>BS/FL</td>
<td>Scientific concept change and reasoning</td>
</tr>
<tr>
<td>Laio &amp; She (2009)</td>
<td>8</td>
<td>PCs-web-based</td>
<td>I/I</td>
<td>Creative thinking</td>
</tr>
<tr>
<td>Liu (1998)</td>
<td>3-4</td>
<td>PCs-Hypermedia authoring</td>
<td>PBL</td>
<td>Computer attitude, achievement</td>
</tr>
<tr>
<td>Liu et al. (1998)</td>
<td>10-12</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Computer attitude, achievement</td>
</tr>
<tr>
<td>Lynch et al. (2000)</td>
<td>5-6</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Reading literacy</td>
</tr>
<tr>
<td>Lynch et al. (1997)</td>
<td>7-8</td>
<td>PDAs - hand-held texts</td>
<td>I/I</td>
<td>Oracy</td>
</tr>
<tr>
<td>Macaruso &amp; Walker (2008)</td>
<td>K</td>
<td>PCs - Multimedia</td>
<td>BS/FL</td>
<td>Literacy skills</td>
</tr>
<tr>
<td>Matthew (1997)</td>
<td>3</td>
<td>PCs - Multimedia</td>
<td>BS/FL</td>
<td>Reading</td>
</tr>
<tr>
<td>McDonald &amp; Hannafin, (2003)</td>
<td>3</td>
<td>PCs - games</td>
<td>BS/FL</td>
<td>Social studies</td>
</tr>
<tr>
<td>McNamara et al. (2006)</td>
<td>8-9</td>
<td>PCs - Interactive</td>
<td>BS/FL</td>
<td>Reading</td>
</tr>
<tr>
<td>Michael (2001)</td>
<td>7</td>
<td>PCs-computer simulation</td>
<td>PBL</td>
<td>Product creativity</td>
</tr>
<tr>
<td>Mitchell &amp; Fox (2001)</td>
<td>K</td>
<td>PCs - Multimedia</td>
<td>BS/FL</td>
<td>Phonological awareness</td>
</tr>
<tr>
<td>Nicolau et al. (2007)</td>
<td>4</td>
<td>Computer labs</td>
<td>I/I</td>
<td>Graphic interpretation</td>
</tr>
<tr>
<td>Raghavan et al. (1997)</td>
<td>6</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Geometry</td>
</tr>
<tr>
<td>Roberts &amp; Stephens (1999)</td>
<td>9</td>
<td>PCs- Interactive</td>
<td>PS</td>
<td>Geometry</td>
</tr>
<tr>
<td>Authors</td>
<td>Quantity</td>
<td>Type(s)</td>
<td>BS/FL</td>
<td>Subject(s)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------</td>
<td>--------------------------------</td>
<td>--------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Ross et al. (2001)</td>
<td>1-2</td>
<td>PCs- Interactive</td>
<td>BS/FL</td>
<td>Computer skills, self-efficacy</td>
</tr>
<tr>
<td>Ross et al. (2009)</td>
<td>7-10</td>
<td>PCs- Interactive web-based tools</td>
<td>PS</td>
<td>fraction</td>
</tr>
<tr>
<td>Rotbain et al. (2008)</td>
<td>10-12</td>
<td>PCs-animations</td>
<td>I/I</td>
<td>biology</td>
</tr>
<tr>
<td>Shemer (1998)</td>
<td>7-12</td>
<td>PCs- simulation games</td>
<td>BS/FL</td>
<td>moral development</td>
</tr>
<tr>
<td>Thomas &amp; Hofmeister (2002)</td>
<td>3 - 4</td>
<td>PCs-message board</td>
<td>-</td>
<td>Literacy</td>
</tr>
<tr>
<td>Tsou et al. (2002)</td>
<td>6</td>
<td>PCs -Web-based multimedia</td>
<td>BS/FL</td>
<td>EFL</td>
</tr>
<tr>
<td>Tusei (2011)</td>
<td>4</td>
<td>PCs-collaborative learning environment</td>
<td>BS/FL</td>
<td>Reading, self-concept</td>
</tr>
<tr>
<td>Waxman &amp; Huang (1997)</td>
<td>6 &amp; 8</td>
<td>PCs - Level of tech use</td>
<td>BS/FL</td>
<td>Motivation, anxiety</td>
</tr>
<tr>
<td>Weiss et al. (2006)</td>
<td>K</td>
<td>PCs-mediarmia (cooperative)</td>
<td>BS/FL</td>
<td>Mathematics &amp; learning style</td>
</tr>
<tr>
<td>Woodul et al. (2000)</td>
<td>8</td>
<td>PCs-multimedia (cooperative)</td>
<td>PS</td>
<td>Social studies &amp; Self-perception</td>
</tr>
<tr>
<td>Yang &amp; Heh (2007)</td>
<td>10</td>
<td>PCs-web-based virtual lab</td>
<td>BS/FL</td>
<td>Computer attitudes, Physics</td>
</tr>
<tr>
<td>Yang &amp; Tsai (2010)</td>
<td>6</td>
<td>PCs</td>
<td>BS/FL</td>
<td>Number sense, learning attitude</td>
</tr>
<tr>
<td>Ysseldyke et al. (2003).</td>
<td>4 &amp; 5</td>
<td>PCs- learning information system</td>
<td>BS/FL</td>
<td>mathematics</td>
</tr>
</tbody>
</table>
Focus on Form in Live Chats

Wen-Chun Chen* and Zohreh Eslami1

National Chung-Cheng University, 168 University Road, Min-Hsiung, Chia-Yi 62102, Taiwan // 1Texas A&M University, 4232, TAMU, College Station, TX 77843-4232 USA // ginachen@ccu.edu.tw // zeslami@tamu.edu

*Corresponding author

(Submitted July 29, 2011; Revised October 3, 2011; Accepted October 18, 2011)

ABSTRACT

This study investigated the effectiveness of incidental focus on form in promoting second language development in text-based live chats. Sixteen college-level Taiwanese English language learners were partnered with American college students to complete two communicative tasks via synchronous chats on Instant Messenger. Language-related episodes (LREs) were identified and used as a basis for individualized test items in which the participants’ linguistic issues were addressed and recalled. The results of two posttests revealed that focus on form in text-based live chats was highly associated with subsequent second language development. Successful uptake and type of feedback were two major factors for learners’ accurate grammatical and lexical knowledge recall, which serve as pedagogical guidelines to promote noticing in second language development.

Keywords

Focus on form, Task-based, Language development, Live chat, Language-related episodes

Introduction

For input to become intake, a learner is required to focus his/her attention on features of the target language. Long (1991) first coined the term focus on form (FonF) to describe this attentional process. Schmidt (1990) argued that noticing at the level of awareness is the prerequisite of any intake of input. In order to direct learners to notice and attend to target structures, Sharwood Smith (1993) suggested teachers to apply input enhancement techniques to lead learners to identify their linguistic errors through feedback provision. In the current study, we took advantage of the text-based live chats on Instant Messenger and the access to native speakers (NSs) of the target language in an attempt to create a high-level attention-drawing condition for a group of English-as-a-foreign-language (EFL) learners.

Building upon Sharwood Smith’s concept of input enhancement (1993) as a way to draw the learners’ attention to specific linguistic forms, computer-mediated communication (CMC—mainly refers to text-based online interaction in this paper) has been endorsed due to its visual display (Smith, 2003a; Warschauer, 1997). Pertinent studies concluded that CMC allows learners to revisit and analyze incoming messages and craft outgoing messages, when engaged in an intensive interaction. The external input enhancement aims to induce internal awareness. Text-based online chat particularly involves written oral-like conversation and is slower (to various extents) than face-to-face (FTF) interaction, yet retains the authenticity of verbal discourse (Beauvois, 1992). When the two powerful input enhancers—NS’ feedback and text chat—are combined into communicative tasks, a significant amount of incidental FonF shall be anticipated. Nonetheless, incidental FonF (its occurrence and outcome) is unpredictable and individualized, in contrast to planned FonF (Ellis, 2001). Hence, the assessment would be difficult through conventional testing techniques. The purpose of the study thus is to offer an empirical account on how incidental FonF can be examined and elicited through text-based live chats between language learners and NSs. Early studies on FonF were mostly theoretical or descriptive (e.g., Cross, 2002; Ellis, 2001; Ellis, Basturkmen & Loewen, 2001a; Gass, 1997; Schmidt, 1990, 2001). Not until recently did some researchers attempt to quantify the occurrence of FonF in FTF (e.g., Doughty & Williams, 1988; Long, Inagaki, & Ortega, 1998) and online contexts (Loewen, 2005; Shekary & Tahririan, 2006; Williams, 2001), among language learners or between a teacher and students. Nonetheless, how feedback provision from NS peers influences learners’ FonF in CMC has not been studied extensively.

Incidental FonF and Subsequent L2 Development

Although communicative function is prioritized nowadays in Language Education, research shows that meaning-focused instruction may not lead to the development of high-level accuracy (Wesche, 1992). Learners who cannot
simultaneously attend to both linguistic aspects may opt for meaning over form during communication (VanPatten, 1990). Therefore, researchers (e.g., Doughty & Williams, 1998 and Ellis, 2001) have proposed incidental FonF to be a method to integrate meaning-focused and form-focused instruction. Learners alternate their attention between form and meaning.

Incidental FonF can also improve learners’ interlanguage. Swain (1995) claimed that FonF pushes learners to deeper, grammatical processing that potentially has a “significant role in the development of syntax and morphology” (p.138). Given the acquisitional potential of FonF interactions, the question arises as to how a facilitative environment is created in which negotiation of meaning and interactional adjustments can occur; Ellis (2003), Long (1983a, 1983b), and Porter (1983) proposed task-based language learning. Since FonF is likely to occur in a task-based collaborative environment, the evidence and details of its occurrence in relation to subsequent L2 development deserve a comprehensive examination.

Assessing Incidental FonF

Researchers have used “language-related episodes” (LREs) in which the learners either talk about or question language use within the context of carrying out a task in L2 (Swain & Lapkin, 1995, 1998, 2001). LREs are also useful for examining the contribution of FonF activities to L2 learning. Therefore, one suggested method to investigate the relationship between FonF and L2 development is to elicit reproduction of the targeted linguistic items through the use of individualized, tailor-made tests based on the specific problems arising in the LREs during the interaction (Swain, 2001). The examination of LREs can serve as a type of pretest indicating learners’ shortage of specific knowledge (Ellis et al., 2001b). When addressing the difficulties derived from LREs, one can assess the effectiveness of FonF (Ellis, 2001; Swain, 2001). In an LRE, interlocutors notice, discuss, and resolve linguistic issues for meaning negotiation.

Several studies have explored the relationship between FonF and L2 learning by individually testing the linguistic items that occur naturally during a 2-way communication (La Pierre, 1994; Swain & Lapkin, 1998; Loewen, 2005). Of these studies, the one by Loewen (2005) is broader in scope and more refined in research design. He examined the occurrence of FonF in 12 ESL classrooms and administering immediate and delayed posttests to individual learners. The students were able to recall nearly 60% and 50% of the linguistic information respectively in the two tests. His findings indicate the effectiveness of FonF, especially if learners incorporate the targeted linguistic forms into their own production (considered as successful uptake). However, his research relied heavily on teachers’ input in FonF episodes. For large language classes, the interaction between one teacher (the only facilitative force) and many students would decrease the occurrence of FonF. Most importantly in FTF settings, the tension, rush, and spontaneity of oral-aural interactions may reduce learners’ time to think and delay their access to rarely used knowledge. Textual CMC, by contrast, allows learners time to process input and output and thus should be further investigated.

FonF and Text-based Chat

Real-time written conversations retain authenticity of FTF discourse, which may help prepare nonnative speakers (NNSs) for fast-paced oral conversations in L2 (Hudson & Bruckman, 2001). The visual salience of written discourse and the self-paced setting in a text-based medium increase learners’ opportunities to take notice of errors and make output modifications, including self-repairs (Kessler, 2009; Lai & Zhao, 2006; Smith, 2008). Warschauer (1997) asserts that a text-based CMC amplifies learners’ attention to linguistic form. Kitade (2000) claims that English learners have actively engaged in meaning negotiation and collaborative learning during online chats described in her study. Pellettieri (2000) tested the potential of synchronous CMC by analyzing negotiation moves and concluded with positive findings. Lai and Zhao (2006) found that, compared with FTF interaction, online chats enhanced input and promoted awareness of problematic linguistic structures. Shekary and Tahririan (2006) investigated the online negotiations of 16 Iranian English learners and obtained findings consistent with Loewen’s (2005): successful uptake was the strongest predictor of correct responses in the posttests.

However, not all studies support the facilitative role of CMC on FonF as compared to FTF interactions. Iwasaki and Oliver (2003), for example, found less feedback and negotiation of meaning in online contexts. A study by Loewen
and Erlam (2006) on corrective feedback in the chat room revealed that students with lower L2 proficiency paid
limited attention to linguistic form and that successful uptake of negotiated forms was reduced in online interaction.
More recently, Loewen and Reissner (2009) compared incidental FonF in an L2 classroom and two types of chat
rooms (with and without teacher moderation). Their data revealed that FonF was most frequent in the FTF context
and least frequent in the unmoderated online context. The inconclusive results of previous research call for more
studies in this area.

Source of Input in Communicative Tasks

The source of input (interlocutors) is another key factor in L2 development (Gass, 1997; Gass & Varonis, 1985). The
role of NS in NS-NNS online interactions remains underexplored. CMC’s technological capacities give learners
access to NSs of the target language and culture by eliminating the geographic distance. Some studies investigated
the role of NSs in promoting L2 development in NS-NNS interactions (e.g., Long 1983a; Pica, 1988): NSs can easily
juxtapose learners’ errors with target-like forms. In spite of the new trend of World Englishes, scaffolding from NSs
supported by their mastery of L1 is often more stimulating than that from NNSs. For instance, the less-proficient L2
learners in Williams’ (2001) research were more receptive to explicit and immediate feedback by NS teachers than
their more-proficient fellow students. In Lee’s study (2008), the scaffolding from the experts in NS-NNS
collaboration assisted students in reconstructing L2, a process that involved both linguistic and cognitive skills.
While these studies report on NSs’ positive role during L2 learning, other studies have found that language
improvement can also occur through collaboration among learners, especially during CMC (Fernandez-Garcia &
Martinez-Arbelaitz, 2002; Shekary & Tahririan, 2006; Smith, 2003a, 2003b, 2004). Varonis and Gass’ early study
(1985) investigated FTF interactions between NS-NS, NS-NNS, and NNS-NNS dyads. Their results showed that
meaning negotiation occurred more frequently in NNS-NNS dyads than the others. The negotiated interaction in
NNS-NNS dyads provides a “non-threatening forum” (p. 87) to build up language competence and an opportunity to
receive comprehensible input via negotiation. Thus, interlocutors’ impact on L2 learning should be further studied.

Research Questions

The present study employed a quasi-experimental design to examine the occurrence and impact of incidental FonF
by connecting language learners and NS e-pals in communicative tasks via text-based online chat. The following
research questions guided the study:
1. How does incidental FonF (occurrence of LREs) affect learners’ L2 development in a text-based chat setting?
2. What characteristics of LREs best predict L2 development in a text-based chat setting?

Methodology

Participants

The focus of the study was the linguistic progress of 16 college-level Taiwanese EFL learners (two males and 14
females in their first or second year of study) achieved through a series of NNS-NS dyadic live chats. In order to
investigate the effect of NS-NNS task-based CMC on learners’ L2 development, 16 American NS e-pals majoring in
Education were invited to join the project and to partner with the NNSs. Before the experiment began, we
implemented an intermediate-level General English Proficiency Test (a national standardized test similar to TOEFL)
to confirm the homogeneity of the Taiwanese group; students ranged from low-intermediate to high-intermediate. All
participants gave consent to join this research. Week 1–3 was the orientation phase for technological training and
ice-breaking between e-pals. After orientation, each dyad engaged in live chats for approximately 90 minutes per
week (see Table 1).

Decision-making and jigsaw tasks were designed specifically for the participants in the study. The discussion topics
of these tasks were “self-value” and “environmental protection”—relevant issues that youngsters in many nations are
likely to encounter in everyday communication. These tasks were selected on the basis that they would draw
abundant meaning negotiations from interlocutors, raise communicative needs, and increase the likelihood of LREs (Gass, 1997; Pica, 1987).

Table 1. Timeline and stages of the experimental design

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Stages</th>
<th>Data collection (16 participants total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Orientation</td>
<td>LREs in the dyadic correspondence records</td>
</tr>
<tr>
<td>Week 2–3</td>
<td>Ice-breaking activity</td>
<td>Results from the two posttests:</td>
</tr>
<tr>
<td>Week 4–6</td>
<td>Jigsaw task</td>
<td></td>
</tr>
<tr>
<td>Week 7–9</td>
<td>Decision-making task</td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td>Immediate posttest</td>
<td></td>
</tr>
<tr>
<td>Week 14</td>
<td>Delayed posttest</td>
<td></td>
</tr>
</tbody>
</table>

Coding of LREs

LREs were identified from the chat scripts (message archives). In total, 505 LREs were coded for a variety of characteristics based on Loewen’s analysis (2005) (see Table 2). Each LRE consisted of three discourse moves: trigger, response, and uptake (optional), starting with the identification of a language problem and ending with a resolution (Swain 2000, 2001). To estimate the inter-rater reliability of the coding, a sample of 50% of the LREs was coded by another rater familiar with the coding scheme. The kappa coefficients ranged from k=.69 to .95, confirming the validity of the coding.

Table 2. Characteristics of LREs

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Definition</th>
<th>Categories</th>
</tr>
</thead>
</table>
| Type            | When an LRE is instigated | -Reactive: Error correction  
|                 |            | -Preemptive: NNS-initiated query |
| Linguistic focus | Linguistic target | -Grammar  
|                 |            | -Vocabulary  
|                 |            | -Spelling |
| Source          | The reason to instigate an LRE | -Code: Inaccurate use of linguistic item with no apparent miscommunication, i.e., form-focused discussions.  
|                 |            | -Message: Problem with understanding meaning, i.e., meaning-focused discussions. |
| Complexity      | Length     | -Simple: Only one response move  
|                 |            | -Complex: More than one response move |
| Directness      | Explicitness of feedback | -Indirect: Implicit (e.g., recast or repetition)  
|                 |            | -Direct: Explicit (e.g., metalingual explanation) |
| Emphasis        | Combination of complexity and directness | -Light: Indirect and simple  
|                 |            | -Heavy: Direct, complex, or both |
| Response        | Type of feedback provided by the NS | -Provision: NS gives information about a language form  
|                 |            | -Elicitation: NS attempts to draw out from NNS a language form or information about a language form |
| Uptake          | NNS response to feedback | -Uptake: NNS produces response  
|                 |            | -No uptake: NNS produces no response |
| Successful uptake | Quality of student response | -Successful uptake: NNS incorporates linguistic information into production or shows solid evidence of understanding  
|                 |            | -Unsuccessful uptake: NNS does not incorporate linguistic information into production |

Note. LRE analysis was adapted from Loewen (2005).

An example of a coding scheme is given in Table 3. The learner misused the word *insane*. The NS sought clarification; the NNS then added more information to make herself understood. Because the learner did not initiate the query, the LRE is reactive. The LRE focused on the word insane, the linguistic focus is vocabulary. The meaning of a problematic linguistic item affected message comprehension; it was categorized as a meaning-focused LRE. The LRE contained more than one response move (turn-taking); it was classified as complex. The NS offered an explicit
explanation, making it a direct LRE. Because this LRE is both complex and direct, a heavy emphasis was assumed. Finally, the NNS acknowledged and incorporated linguistic information provided in the response into the production. Hence, it was considered an instance of both uptake and successful uptake.

<table>
<thead>
<tr>
<th>Jolie (NNS): He keeps getting insane. … I just couldn't bear it anymore.</th>
<th>Characteristics</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary (NS): What do you mean by insane?</td>
<td>Type</td>
<td>Reactive</td>
</tr>
<tr>
<td>Jolie: So I just threw all his clothes that he put in my room back to his room furiously.</td>
<td>Linguistic focus</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>Mary: Haha…did he learn his lesson?</td>
<td>Source</td>
<td>Meaning</td>
</tr>
<tr>
<td>Jolie: Oh, he got very bad manner. He's mean definitely. And he likes to take my things without asking me first. He's always mean and stuck-up.</td>
<td>Complexity</td>
<td>Complex</td>
</tr>
<tr>
<td>Mary: Oh, I thought he is crazy or something. Maybe rude would be a better word to describe your brother.</td>
<td>Directness</td>
<td>Direct</td>
</tr>
<tr>
<td>Jolie: Got it. Yes, he is very rude to me and my parents all the time.</td>
<td>Emphasis</td>
<td>Heavy</td>
</tr>
</tbody>
</table>

The matching suppliance test item:

I don’t like my boss; he is really _______. He talks down to us and yells at us all the time.

### Posttests and Scoring of Test Responses

After the 8-week NS-NNS correspondence, a test was administered to the NNSs in the following week (immediate test) and in week 14 (delayed test) to elicit NNSs’ FonF memories. Both tests were tailor-made for each NNS based on the items each dyad had discussed in LREs. The 1-month interval between the two posttests was set to examine if NSs’ input became NNSs’ intake and was stored in short-term and long-term memories, reflecting the L2 acquisition process; as Gass (1997) postulated: “input→intake→delay or incubation period → integration” (p. 6). Not substituting longitudinal research, this study attempted to fill the gap in the literature and to propose empirical evidence of the subsequent learning through live chats within the given timeframe. Posttests served as a quantitative index of learners’ L2 development through FonF.

The test formats and the scoring procedures were adapted from Loewen (2005). Three test formats were used: suppliance (i.e., fill in the blank), correction, and spelling. A suppliance test was used primarily for the meaning-focused LREs (see the example in Table 3). For the correction test items, the NNSs were required to correct sentences they had erroneously produced during their interactions, i.e., form-focused LREs. For the spelling test items, NNSs were asked to identify the correct spelling of the words that appeared in the LREs. The answers to the test items were coded into three categories: (1) correct (the response matched the targeted item or mentioned synonyms in the tested LRE), (2) partially correct (the response on the targeted linguistic features was improved in some way but not entirely accurate), and (3) incorrect (the NNS failed to reproduce the targeted item in the LRE). The examples of LRE 1, 2, and 3 as well as the matching test items are from the actual data without modifications.

**LRE 1.** Correct test response:

Jackie (NS): …Women used to wear items called corsets…in order to have an hour glass shape. Have you heard of this term before?

Sara (NNS): I’ve seen that before. But I don’t know why they did that.

Jackie: They wanted to impress the men … I can’t imagine wearing one myself. I don’t do well with pain!

Sara: What is hour glass shape?

Jackie: hour glass shape is when women try to make their waists as tiny as possible and then their hips are curvy and larger.

Sara: that must be uncomfortable…

Test Item (suppliance):

What is an hour glass shape when people describe women?

Correct test response (the linguistic issue was accurately recalled):

*Women with tiny waists and curvy and larger hips.*
LRE 2. Partially correct response:
Dan (NNS): … there are quite a few people looking him as an environmental warrior.
Jamie (NS): looking at him?
Jamie: Yes he is.
Dan: You mean I should use at not as?
Jamie: No it should read “I think there are quite a few people looking at him as an environmental warrior.”
Dan: Oh, I see.
Test Item (correction):
Find an error in the following sentence.
Why are people looking me today? Did I dress funny?
Partially correct answer (the answer did not reflect the issue discussed in the LRE but was grammatically and semantically acceptable):
Why are people looking for me? Do I dress funny?

LRE 3. Incorrect test response:
Jackie (NS): Even though I may never be famous someday…and that is all I can ask for.
Sara (NNS): I certain that. And people around us are also important to us
Jackie: What do you mean by “I certain that?” Do you mean you are sure, you are certain of that?
Sara: Yes.
Test Item (correction):
There is an error in the following sentences. Please correct it.
A: I think it is important that we respect ourselves.
B: I certain that.
Incorrect test response: I certain that. (The respondent repeated the same grammatical error discussed in the LRE.)

To avoid a duplication of test items, the immediate test focused on the first half of the experiment and the delayed test covered the second half. A total of 425 LREs were tested (84 % of the total LREs produced). The LREs in which the NS-NNS dyads could not reach resolution on their linguistic issues were omitted (16%).

Data Analysis

Following the identification of LREs and their characteristics, frequency counts were used to present an overview of the generated LREs. The first research question targets the possible impact of FonF on subsequent L2 development (i.e., posttest performance). Therefore, a chi-square analysis was performed, with the significance level set at $\alpha = .05$.

To determine which characteristics of LREs best predict L2 development (test scores), logistic regression analyses were used (Table 4, adapted from Loewen, 2005). Three logistic regressions were performed for three categories (correction, suppliance, and spelling) to reflect three linguistic foci (grammar, vocabulary, and spelling). Each independent variable was added stepwise to the regression models. Each step added the variable that resulted in the greatest change to the model. When a variable did not significantly contribute to the model, it was excluded. For this regression analysis, a significance level of $\alpha=.15$ was chosen because a level of $\alpha=.05$ is too stringent and might exclude important variables (Hosmer & Lemeshow, 2000). All inferential statistics were performed using SPSS.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value=0</th>
<th>Value=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test score</td>
<td>Incorrect</td>
<td>Correct</td>
</tr>
<tr>
<td>Type</td>
<td>Reactive</td>
<td>Preemptive</td>
</tr>
<tr>
<td>Linguistic focus</td>
<td>Grammar</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>Source</td>
<td>Code (Form)</td>
<td>Message (Meaning)</td>
</tr>
<tr>
<td>Complexity</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Directness</td>
<td>Direct</td>
<td>Indirect</td>
</tr>
<tr>
<td>Emphasis</td>
<td>Light</td>
<td>Heavy</td>
</tr>
<tr>
<td>Response</td>
<td>Provision</td>
<td>Elicitation</td>
</tr>
<tr>
<td>Uptake</td>
<td>No uptake</td>
<td>Uptake</td>
</tr>
<tr>
<td>Successful uptake</td>
<td>Unsuccessful uptake</td>
<td>Successful uptake</td>
</tr>
</tbody>
</table>

Note. The third linguistic focus Spelling was excluded due to the the available sample size ($n=22$).
Since logistic regression only allows binary data, the dependent variable (the test scores in this study) needed to be dichotomized (incorrect=0, correct=1). Correct and partially correct answers were combined since both types represented a certain level of learning. The output of a logistic regression analysis includes odd ratios (ORs) and 95% confidence intervals for each independent variable. The larger an OR is, the better predictor a particular independent variable is. For a negative OR value (i.e., OR <1.0), the following rule applies: the smaller the OR, the stronger the negative relationship that forms. ORs are also often used as effect size measures in binary logistic regression (Ott & Longnecker, 2001).

Because the reliability of the individualized testing could not be obtained through conventional test formats, construct validity was chosen to ensure the suitability of the test items designed for the present study. The aim was to verify that the test items actually measured a learner’s ability to reproduce or recall the linguistic knowledge generated in the LREs. The test items and the corresponding LREs were reviewed independently by two Taiwanese teachers to confirm validity. Each assessed the items and categorized them into three categories: appropriate, inappropriate, and uncertain. The agreement between the two raters was 97%. The debatable test items (3%) were withdrawn from the posttests.

### Results and Discussion

In this section, the findings are reported and discussed in an integrative manner by the order of the two research inquiries. As shown in Table 5, the 16 NS-NNS dyads produced a total of 512 LREs ($\bar{x}$=32, $SD$=13.2). The quantity of LREs implied that NNS participants were comfortable enough to ask many questions and freely express their thoughts, despite their limited L2 proficiency. They were not concerned about the awkwardness that often exists in FTF classrooms when students seek clarification from the teacher or their classmates. NSs’ role of target language expert induced learners’ FonF (Loewen & Reiseener, 2009), without creating the teacher-student power distance. Considering the findings of the current research and the aforementioned studies, FonF obviously sustains across different contexts: NS-NNS peer interaction, NS teacher-NNS student interaction, FTF setting, and CMC. Noteworthy is the balanced number of grammar-related and vocabulary-related LREs (257 vs. 255) in this study, suggesting that the FonF in the communicative tasks successfully integrated vocabulary (meaning-focused) and grammar (form-focused) learning.

<table>
<thead>
<tr>
<th>Dyads</th>
<th>Total LREs</th>
<th>Total number of words of LREs</th>
<th>Total LREs tested</th>
<th>Percentage of LREs tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyad 1</td>
<td>23</td>
<td>1497</td>
<td>19</td>
<td>83%</td>
</tr>
<tr>
<td>Dyad 2</td>
<td>20</td>
<td>1100</td>
<td>19</td>
<td>95%</td>
</tr>
<tr>
<td>Dyad 3</td>
<td>30</td>
<td>1478</td>
<td>28</td>
<td>93%</td>
</tr>
<tr>
<td>Dyad 4</td>
<td>50</td>
<td>3645</td>
<td>44</td>
<td>88%</td>
</tr>
<tr>
<td>Dyad 5</td>
<td>29</td>
<td>2208</td>
<td>22</td>
<td>76%</td>
</tr>
<tr>
<td>Dyad 6</td>
<td>22</td>
<td>1655</td>
<td>19</td>
<td>86%</td>
</tr>
<tr>
<td>Dyad 7</td>
<td>19</td>
<td>1284</td>
<td>19</td>
<td>100%</td>
</tr>
<tr>
<td>Dyad 8</td>
<td>19</td>
<td>1284</td>
<td>19</td>
<td>100%</td>
</tr>
<tr>
<td>Dyad 9</td>
<td>34</td>
<td>2474</td>
<td>25</td>
<td>74%</td>
</tr>
<tr>
<td>Dyad 10</td>
<td>29</td>
<td>1620</td>
<td>28</td>
<td>96%</td>
</tr>
<tr>
<td>Dyad 11</td>
<td>23</td>
<td>871</td>
<td>21</td>
<td>91%</td>
</tr>
<tr>
<td>Dyad 12</td>
<td>26</td>
<td>1804</td>
<td>26</td>
<td>100%</td>
</tr>
<tr>
<td>Dyad 13</td>
<td>29</td>
<td>2417</td>
<td>22</td>
<td>76%</td>
</tr>
<tr>
<td>Dyad 14</td>
<td>38</td>
<td>2621</td>
<td>34</td>
<td>89%</td>
</tr>
<tr>
<td>Dyad 15</td>
<td>26</td>
<td>1632</td>
<td>23</td>
<td>88%</td>
</tr>
<tr>
<td>Dyad 16</td>
<td>44</td>
<td>1110</td>
<td>30</td>
<td>68%</td>
</tr>
<tr>
<td>Total</td>
<td>512</td>
<td>32426</td>
<td>425</td>
<td>86%</td>
</tr>
</tbody>
</table>

To address the first research question regarding NNSs’ performance in posttests, a total of 425 LREs were analyzed. The descriptive statistics in Table 6 indicate that participants retained a substantial amount of knowledge gained from feedback in FonF, which possibly contributed to subsequent L2 development. Participants were able to correctly recall and reproduce 70.00% of the target linguistic items in immediate test and 69.39% in the delayed test. Partially correct responses were about the same between the two posttests (7.22% vs. 8.16%). Finally, the incorrect
answer rates were 22.78% and 22.45% in the immediate and delayed tests. A chi-square analysis was performed to find any possible significant differences in the distribution of correct responses between the two posttests ($X^2 (2, n=425) =.318, p>.05$). There was no significant decrease in knowledge retention over the 1-month period. The result remained positive as that in La Pierre (1994), Swain and Lapkin (1998), Williams (2001), Loewen (2005), and Shekary and Tahririan (2006).

### Table 6. Test results

<table>
<thead>
<tr>
<th>Test Responses</th>
<th>Immediate</th>
<th></th>
<th>Delayed</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Correct</td>
<td>126</td>
<td>70.00</td>
<td>170</td>
<td>69.39</td>
<td>296</td>
<td>69.65</td>
</tr>
<tr>
<td>Partially Correct</td>
<td>13</td>
<td>7.22</td>
<td>20</td>
<td>8.16</td>
<td>33</td>
<td>7.76</td>
</tr>
<tr>
<td>Incorrect</td>
<td>41</td>
<td>22.78</td>
<td>55</td>
<td>22.45</td>
<td>96</td>
<td>22.59</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>245</td>
<td>425</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the current research, learners received immediate feedback from their NS peers. Given the linguistic authority status and the curiosity of the never-met NS e-pal from the target culture, the intermediate level learners took e-pals' feedback seriously. The e-pal partnership permitted NS e-pals to act in part as teachers and in part as peers. Mastery of L1 allows NSs to apply various strategies and tactics to negotiate meaning with NNSs (Long, 1983a; 1983b). Nonetheless, the outcome might be different if the feedback from NSs was presented in an oral-aural (FTF) form: the immediacy might reduce the explicitness due to anxiety and haste, especially for L2 learners with low confidence or proficiency (Williams, 2001). Textually displayed chat scripts could also reduce NNSs’ cognitive burden. Communication medium influences learning, and so does input source. If the e-pals were NNSs, the intermediate level participants might not be as receptive in LREs (Williams, 2001). Although collaboration with NNSs from other cultural or linguistic backgrounds may still stimulate intercultural curiosity and linguistic scaffolding, the expert-level feedback is often unattainable (Lee, 2008). Chapelle (1998) and Gass (1997) asserted that the involvement of NSs facilitates and reinforces a learner’s apperception, comprehension, intake, and integration, which could lead to NNSs’ long-term memory retention.

The second research question asked which LRE characteristics best predict subsequent L2 development; three binary logistic regression analyses with dependent (posttest scores) and independent (LRE characteristics) variables were performed. In the logistic regressions, the most significant variable(s) with the highest ORs served as the strongest predictor(s) of NNSs’ L2 improvement. All significant predictors ($p <.05$) in the respective models (overall, correction, and suppliance test types) are listed in Table 7. The most significant predictors with the highest ORs in each model are marked with an asterisk (*). Note that the spelling test type was excluded since the available sample size ($n=22$) was too small to make any claims for this type.

### Table 7. Predictors in three logistic regression models

<table>
<thead>
<tr>
<th>Test item type</th>
<th>Predictors</th>
<th>OR</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>Response</td>
<td>1.166</td>
<td>3.312</td>
<td>1.965</td>
</tr>
<tr>
<td></td>
<td>Successful uptake</td>
<td>1.612</td>
<td>4.791</td>
<td>2.779*</td>
</tr>
<tr>
<td></td>
<td>Directness</td>
<td>.314</td>
<td>.806</td>
<td>.503 (1.988)</td>
</tr>
<tr>
<td>Correction</td>
<td>Response</td>
<td>1.100</td>
<td>4.061</td>
<td>2.114</td>
</tr>
<tr>
<td></td>
<td>Successful uptake</td>
<td>1.096</td>
<td>4.420</td>
<td>2.201*</td>
</tr>
<tr>
<td>Suppliance</td>
<td>Source</td>
<td>1.366</td>
<td>32.963</td>
<td>6.711*</td>
</tr>
<tr>
<td></td>
<td>Directness</td>
<td>.061</td>
<td>.665</td>
<td>.202 (4.950)</td>
</tr>
<tr>
<td></td>
<td>Successful uptake</td>
<td>1.883</td>
<td>11.869</td>
<td>4.728</td>
</tr>
</tbody>
</table>

Note. For predictors with ORs less than 1 (when y=0), the reciprocal values are given in parentheses. $\alpha=.05$ was chosen as the cut-off point to exclude less significant predictors.

The first regression analysis evaluated the correction test items (mainly for grammar-related/form-focused LREs). Two significant variables were shown in the model (see Table 7). Successful uptake, with the highest OR value of 2.2, suggested that correct responses were twice more likely when the LRE contained successful uptake rather than unsuccessful uptake. The second powerful predictor was elicitation of response ($OR=2.1$); elicitation of response was twice more likely to induce a correct test response from the learners. These results correspond to Loewen’s (2004). NSs’ elicitations redirected NNSs’ attention to grammatical structures used in the context of learning tasks, leading
them to consciously perceive the underlying patterns involved, helping them to locate the source of the problem, and engaging them in self-corrections (Shaffer, 1989). Supported with the textual display in live chats, NSs’ frequent use of communication strategies raise NNSs’ level of attention and enhance the linguistic input (Sharwood Smith, 1993).

In the second regression, the suppliance test (mainly for vocabulary-related/meaning-focused LREs) was analyzed. Two predictor variables, source (OR = 6.7) and successful uptake (OR = 4.7), were significant in the model (see Table 7). The former signified that learners correctly recalled the LREs addressing message comprehension almost seven times more than form-focused ones. The second predictor indicates that successful uptake induced almost five times more correct test responses than uptake. Successful uptake was significant in all three models as a strong predictor of correct test responses. These findings are in accordance with the studies by Ellis et al. (2001a), Loewen (2004, 2005), and Shekary and Tahririan (2006): successful uptake is a notably strong predictor of subsequent L2 development.

In addition to successful uptake, elicitation (OR = 1.65) was strong, too. As shown in the excerpts of LRE 1 and 3, NS e-pals employed elicitation as a response strategy to induce pushed output from NNSs (e.g., a confirmation check). The learners reconsidered their problematic expressions and made an attempt to retrieve less-used knowledge (also see LRE 4). This finding corroborates the result of Lee (2008) with regard to NSs’ scaffolding effect on NNSs’ interlanguage quality. Lyster (2001) referred elicitation as a way to encourage participation in meaning negotiation. A text chat environment is more favorable in this regard than FTF classes: each NS-NNS dyad has sufficient time negotiating for detailed meaning or form in comparison.

**LRE 4. Elicitation response:**
Katie (NNS): I think I am a good person because I work for myself to keep the life go on.
Dolly (NS): What do you mean by, "...I work for myself to keep the life go on,"?
Katie: I mean I make my own money.
Dolly: Aha, I see!

The final significant predictor in the overall regression model was directness (i.e., explicit feedback, OR = 1.99). Correct responses were almost twice more likely to occur when the NSs gave NNSs direct feedback. Ellis, Basturkmen, and Loewen (2002) and Ellis, Loewen and Erlam (2006) all emphasized the importance of the frequent use of clear signals or instruction to raise learners’ awareness to the core problems. In LRE 5 the NNS tested her assumption about the meaning of unfamiliar verbs. The self-regulated learning hence occurred and was reinforced by her NS e-pal. The timely clarification and feedback provision, with a meaningful example and comprehensible explanation of the nuances between the similar verbs, were undoubtedly critical for NNS’ intake. Although teachers in classrooms use implicit feedback (e.g., recast or repetition) more often, “students may fail to notice the difference between his/her utterances and the recast. This is because the corrective function of a recast is not always apparent” (Ellis et al., 2002, p.425). Less proficient learners might not have reached the readiness to respond to an implicit input (Williams, 2001). The buffered interaction via online chat creates a non-threatening environment for direct corrections or clarifications. NNSs are less pressured or frustrated than how they might be in a conventional classroom. The fact that these intermediate level learners were able to chat with an NS in English everyday was already empowering to them, considering that NS instructors were unavailable at that particular school.

**LRE 5. Direct feedback:**
Katie (NNS): It is not so good to use the plastic bags. When it burns, it may destroy the air.
Katie: How do you say that? Damage or destroy?
Dolly (NS): I would say damage.
Katie: Oh! I got it.
Dolly: Or destroy. Either one will work there. They mean different things. It depends on what you want to say. Damage means hurt or mess up. Destroy means it is no longer there.
Katie: I didn’t know that.

In summary, the results of both immediate and delayed posttests revealed that LREs in online context were associated with L2 development. Successful uptake and type of feedback were two major factors for learners’ accurate grammatical and lexical knowledge recall.
Conclusion and Implications

Based on the findings of the study, two input enhancers, feedback from NS e-pals and text-based online chat, facilitated learners’ FonF and subsequent L2 development. Learners effectively retained the linguistic issues discussed in LREs, including proper use of grammar and vocabulary. The results further suggest that successful uptake and the type of feedback (elicitations and explicit responses) were two major factors for NNSs’ accurate recall. The results also indicate that the quality of uptake was more crucial than its presence. These elements should be taken into consideration during teacher-fronted classes or peer work in CMC and FTF contexts.

Taiwanese, like the many Asian language learners from a collectivist cultural background, tend to be more reserved (Hofstede, 1986; Katchen, 2002). Error corrections by either the teacher or classmates can be stressful or intimidating in a conventional classroom setting. Live chat between e-pals offers learners an informal and fun social environment that fosters immersive learning. The NNS participants in the present study were highly receptive to the corrective feedback from their NS e-pals. This result could be related to a reduced power distance between NS-NNS peers as compared to that between a teacher and students, or to the trust in the feedback from NS e-pals as well as the textual display of CMC. The pedagogical implication of the study is not to highlight the linguistic mastery of NSs, but instead to underscore the significance of creating an attention-raising environment, inducing various FonF in a natural interaction, and fostering prompt and quality feedback provision to convert input into intake. The limitations of the present study lie with the given timeframe and the scale of the experiment. Future research of FonF in synchronous CMC might address the issues of memory retention rate in a longitudinal work or FonF in aural-oral online interaction as a parallel study.

References


Effect of a Virtual Chemistry Laboratory on Students’ Achievement

Zeynep Tatli and Alipasa Ayas

Karadeniz Technical University Faculty of Education, Trabzon, Turkey // 1Bilkent University, Faculty of Education, Ankara, Turkey // zeynepktu@hotmail.com // alipasaayas@yahoo.com

(Submitted June 10, 2011; Revised October 6, 2011; Accepted December 28, 2011)

ABSTRACT

It is well known that laboratory applications are of significant importance in chemistry education. However, laboratory applications have generally been neglected in recent educational environments for a variety of reasons. In order to address this gap, this study examined the effect of a virtual chemistry laboratory (VCL) on student achievement among 90 students from three different ninth-grade classrooms (an experimental group and two control groups). Study data were gathered with pre and post chemical-changes unit achievement (CCUA) Test, laboratory equipment test (LET), and unstructured observations. The collected data were analyzed using SPSS (version 16.0). Comparisons were made within and between groups. It was concluded that the developed virtual chemistry laboratory software is at least as effective as the real laboratory, both in terms of student achievement in the unit and students’ ability to recognize laboratory equipment.

Keywords

Virtual laboratory, Chemistry laboratory, Students’ achievement, Learning environment

Introduction

Chemistry is perceived by students as a challenging subject, since it is difficult to construct the abstract concepts frequently encountered in the subject area (Ayas & Demirbas, 1997; Nakleh, 1992). Although Turkish students study chemistry as a minor part of the primary-school science course, it is first encountered as a separate course during the ninth grade. More than 70% of these students took the course for the first and last time (Ministry of National Education 2007). Therefore, achievement in the chemistry course during this period profoundly influences students’ branch preferences in their subsequent education.

Previous studies of ninth-grade chemistry topics found that students can understand the course unit on physical and chemical changes (Ayas & Demirbas, 1997), but have difficulty understanding events at the micro level and explaining chemical changes in relation to chemical bonds (Mirzalar Kabapinar & Adik, 2005). In addition, the literature shows that students have difficulty in constructing the topic of the chemical-changes unit in their minds; and that teachers do not support students adequately during this construction process (Palmer & Tregast, 1996; Ayas & Demirbas, 1997; Ayas, Karamustafaoglu, Sevim & Karamustafaoglu, 2002; Kabapinar & Adik, 2005; Ozmen, 2005; Atasoy, Genc, Kadayifci, & Akkus, 2007). The reason for this weakness is frequently attributed to the lack of laboratory practice (Yang & Heh, 2007). The reason for this weakness is frequently attributed to the lack of laboratory practice (Yang & Heh, 2007). Although laboratory work is an indispensable element of understanding chemistry courses, previous studies have reported that it cannot be properly embedded into traditional chemistry courses for various reasons, such as safety concerns, a lack of self-confidence, and an excessive amount of time and effort required to conduct accurate experiments (Elton, 1983; Bryant & Edmunt, 1987; Hofstein & Lunetta, 2004; Durmus & Bayraktar, 2010). Nonetheless, it is not impossible to overcome these obstacles via technology-based alternatives (Okon, Kaliszan, Lawenda, Stoklosa, Rajtar, Meyer, & Stroinski, 2006).

An alternative learning environment, called a virtual laboratory, can help to make this crucial educational application available to students (Kumar Pakala, Ragade, & Wong, 1998; Shin, Yoon, Park & Lee, 2000; Grob, 2002; SAVVIS, 2010; Jeschke, Richter, & Zorn, 2010). Virtual laboratories simulate a real laboratory environment and processes, and are defined as learning environments in which students convert their theoretical knowledge into practical knowledge by conducting experiments (Woodfield, 2005). Virtual laboratories provide students with meaningful virtual experiences and present important concepts, principles, and processes. By means of virtual laboratories, students have the opportunity of repeating any incorrect experiment or to deepen the intended experiences. Moreover, the interactive nature of such teaching methods offers a clear and enjoyable learning environment (Ardac & Akaygun, 2004 Jeschke, Richter, & Zorn, 2010). Table 1 shows a comparison of the reasons why chemistry teachers do not include laboratory applications in their teaching and the solutions offered by virtual laboratories.
Table 1. Problems encountered in chemistry courses and solutions offered by virtual laboratories

<table>
<thead>
<tr>
<th>Reason for teachers’ lack of use of the lab</th>
<th>Alternatives offered by virtual laboratories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety concerns</td>
<td>Experiments that involve risks in the real environment due to poisonous or unsavory gas releases can be safely performed in virtual laboratory environment / uncontrolled explosions (e.g., N\textsubscript{2}) have no real-world consequences, etc.</td>
</tr>
<tr>
<td>Lack of self-confidence</td>
<td>Virtual laboratories help students and teachers with little or no laboratory experience in terms of selecting laboratory equipment, setting up experimental apparatus, and completing the procedure. With the exception of starting the computer or accessing the website hosting the virtual environment software, virtual environments require no prior preparation of laboratory equipment, etc.</td>
</tr>
<tr>
<td>Lack of equipment</td>
<td>As virtual laboratory equipment is not at risk of being broken or lost, users can use virtual laboratories freely. Experiments that cannot be conducted in a real laboratory due to shortages of equipment and materials can be repeated in a virtual lab without any loss.</td>
</tr>
<tr>
<td>Time shortage</td>
<td>Time loss is reduced in virtual laboratories compared to time lost in real laboratories. The experimental procedure in virtual laboratories is similar to that of real laboratories. Understanding and following the experiments is easier in virtual media. After the experiment, it is not necessary to devote time to tidying the virtual laboratory. Students who become accustomed to the virtual laboratory environment can easily repeat the same experiments in the real laboratory environment.</td>
</tr>
<tr>
<td>Weaknesses of confirmation method</td>
<td>The interactive format of the virtual laboratory environment presents the problem case by arousing students’ curiosity. They are made to put forward and test hypotheses, and are also given the opportunity to make generalizations. Since the subsequent experimental steps in the virtual laboratory are pre-planned, based on algorithms, there is no risk of the experiment producing improper results or no results at all. The students are able to research freely within a largely determined framework (Dalgarno, Bishop, Adlong, &amp; Bedgood, 2009; Yu, Brown, &amp; Billet, 2005).</td>
</tr>
</tbody>
</table>

As seen in Table 1, a virtual laboratory may sometimes be a preferable alternative, or simply a supportive learning environment, to real laboratories. A virtual laboratory provides students with opportunities such as enriching their learning experiences; conducting experiments as if they were in real laboratories; and improving their experiment-related skills such as manipulating materials and equipment, collecting data, completing experiment process in an interactive way (with boundless supplies), and preparing experiment reports (Subramanian & Marsic, 2001).

Researchers have determined that instructions carried out with virtual laboratories significantly increase student achievement levels. Virtual environments let students observe the process in more detail, compared to board and chalk activities of the traditional classroom or partially completed experiments of the real laboratory environment. In addition, virtual environments foster attention and motivation towards the course by supporting a discussion platform among partners, peers, and among students and teacher (Hounshell & Hill, 1989; Geban, Askar, & Ozkan, 1992; Kubala, 1998). Furthermore, some researchers even argue that performing experiments within a virtual environment is more effective than performing experiments in real laboratories (McCoy, 1991; Geban et al., 1992; Svec & Anderson, 1995; Kozma, Chin, Russell, & Marx, 2000; Browne, 2002).

Context of study

According to the literature, the laboratory approach is regarded as an indispensable element of chemistry education, and students subjected to constructivist learning theory-based laboratory instruction exhibit higher achievement scores, deeper attention, and more frequent participation in chemistry course (Duffy & Jonassen, 1991; Aydogdu, 2003; Puacharearn, 2004; Karagiorgi & Symeou, 2005; Celikler, Gunes, & Sendil, 2006; Atasoy, Genc, Kadayifci, & Akkus 2007 Koseoglu & Tumay, 2010) However, it is obvious that learning environments adopting and applying constructivist learning theory should be supported with activities facilitating cooperation and interaction (Ayas & Demirbas, 1997; Baki, 2008), which require more time.

Predict-Observe-Explain (POE) is known to have a positive effect on student learning, and is a recommended strategy to enrich constructivist learning environments (Liew & Treagust, 1998).
However, studies showed that, in traditional learning environments, there are always inconsistencies between student predictions and observations (Champagne, Klopf, & Anderson, 1980; White & Gunstone, 1992; Liew & Treagust, 1998; Kearney & Treagust, 2001. Such environments also make students reserved and cause them to refrain from expressing their opinions directly (Sheppard, 2006). In contrast, virtual learning environments enable learners to repeat the events several times without hesitation, to zoom in and out, and to watch in slow motion being questioned in any way (Kearney & Treagust, 2001.

The present study evaluated how student achievement levels are affected by the use of virtual chemistry laboratory (VCL) software. The software does the following:

- Models the positive influences on student achievement mentioned above.
- Eliminates adverse effects of virtual laboratories.
- Increases student participation.
- Visualizes macro, micro, and symbolical level presentations of the experiment.
- Provides a strategy that follows the procedural steps of the POE strategy.

**Purpose**

The principal aim of this study is to determine the effect of a virtual chemistry laboratory on student achievement in course units entitled “chemical changes” and “recognizing laboratory materials and equipment.” The course units form part of the ninth-grade chemistry curriculum, and the virtual learning environment approach is based on constructivist learning theory, following the steps of POE strategy.

**Material and method**

The following steps were used to develop the virtual chemistry laboratory. First, topics and units in which students experience difficulties were determined by interviews with chemistry teachers ($n = 20$) instructing ninth-graders. These data were supported by the literature and led to the selection of the chemical changes unit, which constitutes 20% of the ninth-grade chemistry curriculum (Ministry of National Education, 2007).

We then conducted a pilot study in which the experiments within the chemical changes unit were performed by student teachers in the General Chemistry I course at the Karadeniz Technical University. The experimental processes were recorded, and the records were used to evaluate the reality of the experiments, which would be transferred to the virtual environment, and to determine what difficulties the student teachers encountered. We reviewed the virtual laboratory applications in the literature.

We prepared the content of the virtual chemistry laboratory software. Before the preparation, previous virtual laboratory applications and positive and constructive comments directed to these applications were reviewed. Following this review, it was determined that previous virtual laboratory applications used in the chemical changes unit contained only one or two dimensions. That is, some of the previous applications focused on visual dimensions but no instructional model was applied during development, whereas other applications prioritized molecular level representations but neglected student interaction. Still some others presented a very large laboratory environment but users were found to lose time, get tired, and, if they eventually managed to use these environments, took longer to arrive at solutions (Tatli, 2011).

We considered these findings during the development of the VCL used in the current study. We developed the virtual learning environment in the following stages:

- Determined the computer programs and applications to be used in the software (3D Max, Adobe Photoshop, Macromedia Flash 8, Adobe Audition, and CrazyTalk)
- Developed software content
- Modeled the laboratory interface, material and supply cupboards, experimental materials and equipment and avatars necessary for the virtual chemistry laboratory
- Created and recorded scripts for avatars and the software’s scientific content
- Developed the virtual chemistry laboratory
Collected expert opinions about the virtual laboratory
Finalized the virtual chemistry laboratory before the pilot application
Chose the school for the pilot application and provided administrative permissions
Piloted the virtual chemistry laboratory

The pilot scheme let us see the shortcomings of our virtual chemistry laboratory that we piloted. By taking into consideration the results we got from the surveys we improved and finalized the virtual chemistry laboratory before the main application. Some screenshots of the software at that stage are given in Figure 1.

Figure 1. Screenshots from the virtual chemistry laboratory software

The piloted VCL software is more advanced than that of previous studies in terms of the following:
- Introducing experimental rules, laboratory equipment and materials to the user
- Providing the user with a learning environment based on constructivist learning theory and POE strategy
- User independency across a wide range of applications and parameters
- The recording of pre-information about the user for comparison with final results
- Presenting macro, micro and symbolical dimensions of the experiment simultaneously
- Providing the user with information about the relationship between the experiment and real life
- Presenting experiments in video format and directing some inquiries to the user, simultaneously presenting synchronized sound, music, visual and information (virtual TV component)
Sample

The sample group of this study consisted of one experienced chemistry teacher and 90 students from one high school. Table 2 presents the total duration saved for the chemical changes unit and the classroom/laboratory.

<table>
<thead>
<tr>
<th>Group name</th>
<th>Preferred experiment environment</th>
<th>Class</th>
<th>Laboratory</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EG</td>
<td>VCL</td>
<td>4</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>CG-I</td>
<td>Real laboratory</td>
<td>9</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>CG-II</td>
<td>Real laboratory</td>
<td>8</td>
<td>3</td>
<td>11</td>
</tr>
</tbody>
</table>

To check the quality and functions of the developed materials, in-depth interviews were carried out with the participating teacher. Participating students were randomly assigned to one of three groups: experimental group (EG, \( n = 30 \)), control-I group (CG-I, \( n = 30 \)) and control-II group (CG-II, \( n = 30 \)). The experimental group performed all the experiments in chemical changes unit using the virtual chemistry laboratory (VCL). The CG-I and CG-II groups were not manipulated, and the teacher taught the unit using their own conventional methods. Only the teacher of the CG-II was encouraged to do all the experiments of the unit in the real laboratory (when it was possible). However, the teacher was not given any additional suggestions and was only informed that one of the researchers would passively observe the session.

Data collection instruments

The data collection tools of the study were: the Chemical Changes Unit Achievement Test (CCUA), the laboratory equipment test (LET), semi-structured interviews, and unstructured observations.

Chemical changes unit achievement test (CCUA)

A validity–reliability study of the CCUA was done by Tatli (2011) to cover all the learning outcomes of the chemical changes unit. In the pilot study, the CCUA test, which initially had 30 items, was applied to 90 ninth-grade students who were formerly instructed in the chemical changes unit. As a result of item analysis, five items with a discrimination index lower than 0.30 were excluded. The item analysis showed that the discrimination index of each item in the test was 0.30 or greater, Spearman Brown reliability coefficient was 0.85, average item difficulty was 0.61, and Pearson product-moment correlation coefficient was 0.74. The pilot study showed that 30 minutes would be sufficient to apply the remaining 25 items.

Laboratory equipment test (LET)

A validity–reliability study of the LET was done by Tatli (2011). The items in the LET, devised by the researchers, were prepared in order to cover all laboratory materials and equipment used in primary school science and ninth-grade chemistry courses. The test was reduced to 28 items in accordance with the opinions of five academics from departments of instructional natural science and chemistry. In addition to these 28 items, a module was added, asking students to enter the names of laboratory materials and equipment into blank spaces beneath color pictures of the related material and equipment. It was decided that 15 minutes would be sufficient to complete the finalized form of the LET.

Procedure

The implementation took six weeks during the spring semester of the 2009–2010 academic year and was conducted at an Anatolian secondary school located in the city center of Trabzon. The control and experimental groups were chosen randomly. However, since within-group manipulation was impossible (i.e., students were registered to related classrooms), the study used a quasi-experimental method. Compared to an experimental method, the quasi-
experimental method is used when it is impossible to choose a completely random sample. The data were collected using the CCUA and LET at the beginning and at the end of the study period as pre-test and post-test measures of student achievement. The implementation process and measurement tools applied are summarized in Table 3.

<table>
<thead>
<tr>
<th>Group name</th>
<th>Data collection tools</th>
<th>Process</th>
<th>Data collection tools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCUA</td>
<td>LET</td>
<td>CCUA</td>
</tr>
<tr>
<td>EG</td>
<td>X</td>
<td>X</td>
<td>*</td>
</tr>
<tr>
<td>CG-I</td>
<td>X</td>
<td>X</td>
<td>**</td>
</tr>
<tr>
<td>CG-II</td>
<td>X</td>
<td>X</td>
<td>***</td>
</tr>
</tbody>
</table>

* The teacher giving instruction to perform all the experiments of the chemical changes unit within the VCL environment. The experiences during the experiments were observed passively and recorded by one of the researchers, using the unstructured observation form.

** The teacher instructed with no manipulation. (The teacher used the real chemistry laboratory for experiments with his/her usual frequency). The teacher’s consent was requested only for observing the lesson. The experiences during the experiments were observed passively and recorded by one of the researchers on the unstructured observation form.

*** The teacher instructed in the real laboratory environment. The experiences during the experiments were observed passively and recorded by one of the researchers on the unstructured observation form.

Data analysis

The SPSS 16.0 statistical analysis program was used to analyze the data. In order to test whether CCUA and LET pre- and post-test scores of CG-I, CG-II, and EG groups were statistically different within group-paired samples t-test, to compare groups one-way ANOVA were carried out (Johnson & Christensen, 2004). Additionally, observation data and interviews were used to support the quantitative data.

Results

Student achievement at chemical-changes unit

Table 3 shows paired samples t-test results, used to check whether there was a significant relationship between pre- and post-CCUA scores of the control and experimental groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>CCUA</th>
<th>N</th>
<th>X'</th>
<th>S</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG-I</td>
<td>Pre-test</td>
<td>30</td>
<td>42.666</td>
<td>12.015</td>
<td>29</td>
<td>2.895</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>30</td>
<td>50.833</td>
<td>12.668</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-II</td>
<td>Pre-test</td>
<td>30</td>
<td>39.666</td>
<td>13.578</td>
<td>29</td>
<td>2.895</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>30</td>
<td>55.333</td>
<td>11.121</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EG</td>
<td>Pre-test</td>
<td>30</td>
<td>39.66</td>
<td>13.578</td>
<td>29</td>
<td>6.388</td>
<td>0.000*</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>30</td>
<td>59.33</td>
<td>11.121</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results show that there were significant differences between the pre- and post-test scores of each group, favoring the post-test (CG-I \[t(29) = -2.895, p < .05\], CG-II \[t(29) = -6.210, p < .05\], EG \[t(29) = -6.388, p < .05\]). When overall pre- and post-CCUA results of the participant students were examined in more detail, it was observed that the average score of the CG-I students increased from 39.66 to 55.33 and CG-II from 42.66 to 50.83, while the average score of EG students increased from 39.66 to 59.33. Student achievement at the end of the chemical-changes unit showed that the greatest increase was achieved in the EG group. This implies that the VCL software supported instruction at least as effectively as the real chemistry laboratory.

Table 4 presents the results of the one-way ANOVA, comparing CCUA pre- and post-test scores of the control and experimental group students.
Table 4. One-way ANOVA results comparing pre- and post-CCUA scores of CG-I, CG-II, and EG students

<table>
<thead>
<tr>
<th>CCUA</th>
<th>Sum of squares</th>
<th>SD</th>
<th>Mean square</th>
<th>F</th>
<th>p</th>
<th>Sig. difference</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Between groups</td>
<td>180.000</td>
<td>2</td>
<td>90.000</td>
<td>0.659</td>
<td>0.520*</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>11880.000</td>
<td>87</td>
<td>136.552</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>12060.000</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td>Between groups</td>
<td>1085.000</td>
<td>2</td>
<td>542.500</td>
<td>4.500</td>
<td>0.014*</td>
<td>EG-CG-I</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>10487.500</td>
<td>87</td>
<td>120.546</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>11572.500</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the results presented in Table 4, there was no significant difference between the groups at the beginning of the study \( F(2-87) = .659, p > .05 \). However, the post-test results showed a significant between-group difference \( F(2-87) = 4.500, p < .05 \). Scheffe’s test was used to determine which groups caused the difference, indicating that the difference was between EG (= 59.33) and CG-I (= 50.83), favoring EG. There were no other significant differences between the groups. This finding suggests that the VCL software was at least as effective as the real chemistry lab when assessed in terms of student achievement in the chemical-changes unit.

**Student success in recognizing laboratory equipment**

A paired samples t-test was used to check whether there was a significant relationship between pre- and post-LET scores of the control and experimental student groups (See Table 5).

<table>
<thead>
<tr>
<th>Group</th>
<th>LET</th>
<th>N</th>
<th>X</th>
<th>S</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG-I</td>
<td>Pre-test</td>
<td>30</td>
<td>24.80</td>
<td>9.65</td>
<td>29</td>
<td>−1.29</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>30</td>
<td>22.19</td>
<td>7.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-II</td>
<td>Pre-test</td>
<td>30</td>
<td>28.53</td>
<td>7.66</td>
<td>29</td>
<td>2.71</td>
<td>.01*</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>30</td>
<td>35.43</td>
<td>12.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EG</td>
<td>Pre-test</td>
<td>30</td>
<td>29.45</td>
<td>9.839</td>
<td>29</td>
<td>13.21</td>
<td>0.00*</td>
</tr>
<tr>
<td></td>
<td>Post-test</td>
<td>30</td>
<td>67.41</td>
<td>13.063</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The post-LET test scores of the CG-I students were lower than the pre-test scores, but this difference was not significant (Table 5) \( t_{29} = 1.299, p > .05 \). However, the pre- and post-test scores of CG-II and EG were significantly different, favoring the post-tests.

In this context, it can be said that the activities conducted by the CG-I group had no effect on students’ recognition of the laboratory equipment, while the activities performed by CG-II and EG students positively affected students’ ability to recognize laboratory equipment.

The results of the one-way ANOVA, comparing LET pre and post-test scores of the control and experimental group students are presented in Table 6.

<table>
<thead>
<tr>
<th>LET</th>
<th>Sum of squares</th>
<th>SD</th>
<th>Mean square</th>
<th>F</th>
<th>p</th>
<th>Sig. difference</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Between groups</td>
<td>363.20</td>
<td>2</td>
<td>181.60</td>
<td>2.19</td>
<td>0.11</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>7213.76</td>
<td>87</td>
<td>82.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7576.97</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td>Between groups</td>
<td>33799.523</td>
<td>2</td>
<td>16899.762</td>
<td>135.56</td>
<td>.000*</td>
<td>EG-CG-I</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>10845.908</td>
<td>87</td>
<td>124.666</td>
<td></td>
<td></td>
<td>EG/C-GII/CG-I</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>44645.431</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td>CG-II/CG-I</td>
</tr>
</tbody>
</table>

*: significant at 0.05
As seen in Table 6, there was no significant difference between the groups at the beginning of the study \( F(2, 87) = 2.19, p > .05 \). However, at the end of the study, there were significant differences between the groups \( F(2, 87) = 135.56, p < .05 \). The Scheffe’s test indicated that there were significant differences between EG and CG-I; EG, and CG-II; CG-II and CG-I.

**Discussion and conclusions**

In previous studies investigating the effects of virtual laboratories on student achievement, virtual laboratory applications induce an expectation of higher student achievement (Yaman, Nerdel, & Bayrhuber, 2008). However, when the previous studies were reviewed, there was no stable relationship between student achievement and the use of virtual laboratories. Some previous studies reported that virtual laboratories positively influenced student achievement (Geban et al., 1992; Burke, Greenbowe, & Windschitl, 1998; Clark, 1998; Monaghan & Clement, 1999; Akpan & Andre, 2000; Jimoyiannis & Komis, 2000; Kennepolh, 2001; Dori & Barak, 2001; Huppert, Lomask & Lazarowitz, 2002; Blaylock & Newman, 2005; Yu et al., 2005; Hughes McLeod, Brown, Maeda & Choi, 2007; Bozkurt, 2008; Limniou, Papadopoulos, Giannakoudakis, Roberts, & Otto, 2007; Chen, 2010), while others reported no significant differences in learning outcomes between traditional environment and virtual environment (Bernard, Abrami, Lou, Borokhovski, Wade & Wozney, 2004; Cavanaugh, Gillan, Kromrey, Hess, & Blomeyer, 2004). Some other studies (Gorghiu, Gorghiu, Alexandrescu, & Borcea, 2009) reported that traditional laboratories were more effective, despite the fact that virtual laboratories provided a variety of benefits.

The results of the present study suggest that virtual laboratories are at least as effective as real laboratories in terms of acquainting students with experiment process (Yu et al., 2005; Dalgarno, Bishop, Adlong, & Bedgood, 2009), providing students with a safe experimental environment (Mercer-Chalmers, Goodfellow, & Price, 2004), allowing students to conduct experiments individually (Bozkurt, 2008), providing users with more options in shorter time with interaction (Regan & Sheppard, 2006; Ozdener & Erdogan, 2001), and simultaneously presenting micro, macro, and symbolic presentation levels to the user (Carlsen & Andre, 1992).

The research team proposes that this result is due to the greater number of experiments that were conducted by the experimental group students compared to the control group students. In addition, some advantages of the virtual laboratory listed below positively affect students’ achievements in the experimental group. These advantages are as follows: the students in the EG focused on the process instead of the materials and equipment; the macro, micro, and symbolic dimensions of the experiments could be investigated in detail with virtual media; students could reach the solution by trying the different choices presented; the software was prepared along with the steps of POE; the software provided students with the opportunity to transfer domain specific knowledge into everyday life.

This conclusion is also supported by the theory that, by maximizing interactivity, virtual laboratory applications render students active thinkers instead of passive observers and thereby construct effective and meaningful learning processes (Trindade, Fiolhai, & Almedia, 2002). It was expected that the adoption of POE to the system increased student-student and student-teacher interactions, reasoning frequency and attentive participation of the students, and made the learning more persistent (Margel, Eylon, & Scherz, 2004; Thomas, Ashton, Austin, Beevers, Edwards, & Milligan, 2004; Karaer, 2007; Tekin, 2008; Chairam, Somsook, & Coll, 2009). The high frequency of abstract concepts that are presented in chemistry (Nakhleh, 1992; Ayas & Demirbas, 1997) is one of the most important factors adversely affecting student achievement (Gabel, 2003; Pekdag, 2010). However, previous studies have stated that virtual laboratories facilitate the formation of conceptual models by providing activities that improve cognitive skills (Kennepolh, 2001; Trindade et al., 2002; Ardac & Akaygun, 2004; Falvo, 2008; Pekdag, 2010). Since chemistry is closely related to daily life (Secken, Morgil, Erokten, Erdem, & Caglayangol, 1999), the use of an “associate with real life” tab within the software is expected to contribute to student achievement.

The pre-LET test scores showed that the achievement levels of the control and experimental groups were quite low and similar to each other. This situation can be attributed to the examination-oriented instruction applied throughout the primary school and, accordingly, the lack of laboratory applications in the study context (Saka, 2002). In Turkey, two highly competitive nationwide exams that place students in upper schooling are held at the end of grades 8 and 12. Since sufficient class hours are not reserved for laboratory applications at primary school level, and because students are not at the heart of the laboratory applications, they cannot recognize laboratory equipment properly (Guzel, 2000). The review of previous studies showed that primary school science courses in Turkey are supported
by little or no laboratory applications. Furthermore, most of the laboratory hours involve only the teachers demonstrating the experiments (Guzel, 2000; Ulucinar, Cansaran, & Karaca, 2004; Ozmen, Demircioglu, & Coll, 2009). Even worse, a study carried out at the university level determined that students cannot properly recognize laboratory materials and equipment (Costu, Ayas, Calik, Ünal, & Karatas, 2005).

When post-LET scores were examined, there was a significant difference in the student achievement scores of the groups. The achievement level of CG-I students dropped 10%, of CG-II students increased 24%, and the scores of EG students increased by 128%. Concerning the effect size, it can be said that 75% of the total variance was the result of the students’ recognition of the laboratory material and equipment. These results indicate that VCL software is important for laboratory activities, and that virtual chemistry laboratories are at least as effective as real laboratories. Previous studies comparing virtual laboratories with real ones also support the finding that virtual laboratories are superior in terms of experiment materials and equipment and they lead to higher student performance, and that virtual laboratories are at least as effective as real laboratories (Gabberd, Hix, & Swan, 1999; Ozdener & Erdogan, 2001; Dalgarno et al., 2009; Ozdener, 2005; Gorghi et al., 2009).

Some experiments can usually only be performed using a demonstration method, for reasons such as a lack of laboratories, insufficient material and crowded classrooms. For these experiments, it is clear that virtual laboratories can provide a valuable alternative to traditional laboratory applications (Ozdener, 2005). Students should be exposed to more laboratory applications and activities so that they can recognize laboratory materials and equipment. Knowing that students who do not have proper pre-knowledge and experience could not be successful while they were doing experiments (Temiz & Kanli, 2005), one can see that a virtual laboratory environment provides students with the opportunity to develop.

At the end of the study, the virtual laboratory software was shown to be at least as effective as real chemistry laboratories. It was determined that students in the control group could complete the experiments with reasonable results; they felt self-confident; they could associate the experiment with daily life; and they had the opportunity to examine macroscopic, molecular and symbolic levels of each experiment. It is anticipated that virtual chemistry laboratories will be adopted as supplementary and supportive elements in future. This will provide not only an effective learning environment but will also minimize school expenditures and the time spent on such activities to a large extent.

References


Ozdener, N., & Erdogan, B. (2001). Improving the virtual laboratories which give the possibility of evaluating the experimental data and giving feedback. Educational Science Journal of MU Faculty of Ataturk Education, 14, 107–120.


Understanding Students’ Competition Preference in Multiple-Mice Supported Classroom

Ben Chang¹, Min-Tun Chuang²* and Spes Ho¹

¹Department of E-Learning Design and Management, National Chiayi University, 85 Wenlong, Mingsuin, Chiayi Hsien 62117, Taiwan // ²Department of Foreign Languages, National Chiayi University, 85 Wenlong, Mingsuin, Chiayi Hsien 62117, Taiwan // ben@mail.ncyu.edu.tw // mtchuang@mail.ncyu.ncu.edu.tw // superspes@gmail.com

(Submitted March 17, 2011; Revised December 15, 2011; Accepted February 22, 2012)

ABSTRACT

Competition has been prevalent in all types of childhood activities including playing, learning, and living. Studies have shown that a well-designed competitive activity can motivate students to perform at an increased level. However, there have been few research studies on the relationship between learners’ preferences and performance in relation to the level of competition in activities, especially in a computer-assisted learning environment. In this study, different levels of group and individual competitions were designed and implemented through an innovative multiple-mice learning system. Sixty 7th grade students were involved to interact with their own group members and compete with other groups to complete English vocabulary tasks. The preferences of students, of different levels of achievement, toward various competition activities were investigated upon finishing the tasks. This paper presents the potential of integrating the multiple-mice learning system into English learning tasks, and most importantly, it reveals the preferences of students regarding activities at various levels of competition in a computer-assisted learning environment.

Keywords

Competition preference, Multiple-mice technology, Learning achievement, Highly interactive classroom

Introduction

The design of social activities and how social activities affect learning have been considered important issues in educational research (Vygotsky, 1978). Among the various social learning activities, competition, an instinctive human behavior, plays a role in the design of a technologically supported learning environment. Obviously, excessively competitive learning activities can have many negative effects. Highly-capable students and less-capable students may perceive the same competitive activity with different feelings (Cheng, Wu, Liao & Chan, 2009). People who walk away from a victory will experience a different emotional state than the ones who walk away from a loss (Kohn, 1992). However, competition is generally considered an effective technique in motivating people to learn and to excel (Yu, Chang, Liu & Chan, 2002; Chang, Yang, Yu & Chan, 2003). Adding an element of competition is widely believed to be a motivation-enriching strategy in play, work, and education (Deci, Betley, Kahle, Abrams & Porac, 1981; Chang, Wang, Peng & Hsu, 2010). Additionally, competition has been suggested as a way to stimulate users’ direct involvement and interest. The use of competition strategies in a classroom is a method that incurs both positive and negative effects. Studies on the challenge of applying competitive strategy in a learning environment therefore should focus on understanding how competition affects the students’ attitudes and how to take advantage of students’ increased motivation and positive learning effects, and decrease the negative effects of competition.

Classrooms are the environment where students spend most of their time and where they encounter learning activities designed to cover cooperation, competition and individual learning. Johnson and Johnson (1998) mentioned that competitions need to be grouped homogeneously. Beyond the explorations of classroom learning activities, researchers are interested in understanding students’ emotional development (Schutz & Pekrun, 2007). Through their understanding of students’ preferences, the more adaptive support we can provide teachers and students. Integrating social activities into the classroom environment is an important issue when designing technology-enhanced classroom learning systems. Digital learning devices in classrooms provide teachers and students with a new interaction media. In the classroom, technology is a facilitator that helps aid quality interaction between teachers and students under a well-designed content and pedagogy (Roschelle & Pea, 2002). In respect to applying technology in classrooms, ensuring students have sufficient access to the necessary technology becomes an important issue.
In technology-enhanced classroom designs, a straightforward idea is to equip each child with one computer, and then they can use the computer at will. Making sure every student has a classroom computer is an important approach that can solve the students’ technology accessibility problem (Chan et al., 2006; Roschelle, Penuel & Abrahamson, 2004). However, the cost of providing every student with one computer is too high for ordinary teachers and students to accommodate. Moreover, classrooms are also the place where students have peers to interact with. Equipping one student with one computer is helpful for learning, but we will argue that letting a group of students interact with their peers on a shared display simultaneously can help break the isolated learning scenario, better attract students’ attention, encourage group work, as well as improve the students’ social interaction. In this particular design, each student is equipped with a thin client, which can be a sensor or a mini device such as a mouse, and each of the students’ sensors (Chang & Chen, 2010) or mini devices are connected via a classroom server. This kind of technology-enhanced classroom design, in comparison to the one-to-one computer concept, is different and is termed as a many-to-one technology-enhanced classroom setting.

The investigation of using multi-user environment such as shared interactive display can be traced back to the 1990s. The system MMM (the multi-device multi-user multi-editor) developed by Bier and Freeman (1991) was the prototype of using multiple-device on a computer. The large shared interactive display enables a group of people to use a computer simultaneously was reported in 1993 (Pedersen et al., 1993), and the multiple-mice technology supported groupware software was available in 1999 (Stewart, Bederson & Druin, 1999). Over the last decade, human-computer interaction researchers have been revising and improving the development of multiple-mice environment in which a computer can be equipped with many mice and the users can use the mice on the computer simultaneously. By using the shared display software, groups of students can be more attractive and participated in learning activities (Scott, Mandyrk & Inkpen, 2003). By equipping each student a mouse, the students can form a group to have face-to-face group learning activity (Infante, Hidalgo, Nussbaum, Alarcon & Gottlieb, 2009). Most of the multiple-mice technology were developed to support face-to-face collaborative activities in the workplace as well as learning space.

In this study, a multiple-mice supported competitive learning environment was designed and implemented for the teachers and students to have face-to-face group competitive activities in the classroom. Solely providing a classroom with a multiple-mice environment is insufficient for practicing social interaction learning activities. Therefore, designing a multiple-mice enhanced classroom should consider not only the technology, but contemplate how to design truly immersive group activities. In this study, a multiple-mice learning system with five different competitive learning activities was used. Through this system, the students can interact with their peers and compete with other groups by using a mouse both individually and simultaneously with other mice on a shared display to complete English vocabulary tasks. The objective of this study is to investigate the students’ preferences toward various competitive activities in this type of multimedia environment. If we can understand the students’ preferences more, we can provide more precise support when designing future activities. Besides, concerning the multiple-mice enhanced classroom design, in comparison to the one-to-one computer environment, we argue that this multiple-mice learning system is much affordable for teachers due to its low cost and is feasible in a regular classroom. In addition, the interactions between students are enhanced by well-designed language tasks.

**MUSCLE: Multiple-Mice Supported Competitive Learning Environment**

Using information technology in classrooms is a trend of educational technology development where students’ accessibility to information technology is a critical issue. No technology-enhanced classroom project can be successful if the students don’t have easy access to the necessary tools. However, the cost of applying information technology in classrooms was previously too high for teachers and students to accommodate. As mentioned above, human-computer interaction researchers have demonstrated the potential of connecting a computer to numerous mice allowing students to simultaneously use the mice with their peers on the same computer.

In this study, by using the multiple-mice technology, a multiple-mice supported competitive learning environment named MUSCLE (Multiple-mice Supported Competitive Learning Environment) was designed. Figure 1 displays a MUSCLE system scenario. In the scenario, the teacher who wants to use the MUSCLE system only needs a projector, a projection screen, a personal computer or notebook, a USB hub, and several general purpose mice. These facilitators are quite common in a regular classroom.
To facilitate students’ performance using MUSCLE, four major functions, Identification & Grouping, System Utility, Individual Competition Mode, and Group Competition Mode, are implemented. Furthermore, Individual Competition Mode is divided into an individual task competition (ITC) activity and an individual rush competition (IRC) activity, and Group Competition Mode is divided into a group coop (GCO) activity, a group competition (GCM) activity, and a group rush competition (GRC) activity. The multiple-mice technology provides an environment where each student has at least one personal cursor to interact with their peers in several competition designed activities. Table 1 lists the functions of MUSCLE system.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification &amp; Grouping</td>
<td>All the usernames are listed on the shared notebook screen. The users identify their names by using their cursors. The name list is edited by the teacher or the student before the activity.</td>
</tr>
<tr>
<td>System Utility</td>
<td>The users can use the system utility to edit the usernames list and the item bank, or configure the parameters of the MUSCLE system.</td>
</tr>
<tr>
<td>Individual Competition Mode</td>
<td>Individual competition mode consists of an individual task competition activity and an individual rush competition activity. Each student in this mode is allocated an individual area on the shared display to complete the activity.</td>
</tr>
<tr>
<td>Group Competition Mode</td>
<td>Group competition mode consists of a group coop activity, a group competition activity and a group rush competition activity. Each group of students answer their questions in a shared area.</td>
</tr>
</tbody>
</table>

(a) Identification & Grouping

Figure 1. Multiple-mice supported classroom scenario

Figure 2. Identification and authoring utility interface
Identification & grouping

In the MUSCLE environment, all the mice cursors can be displayed on a shared notebook, and each student can move their mouse to identify their cursor. Once the students recognize their cursors, the students can click on their names to match their cursors with their names. Figure 2 (a) shows the name list which was edited in advance by the teacher or the students via the editing usernames sub-function of system utility. After completing name assignments, the teacher can enter the next stage of assigning the groups. The teacher can merge students into one group or divide the students into several different groups depending on the needs of the competition modes.

System utility

The system utility has three sub-functions. The first sub-function is editing the username list; the second, editing the item bank; the third, setting system parameters. The teacher or the students can use the username list edit function to edit their usernames. The users can identify their names on the shared display via their cursors. The MUSCLE system provides two question formats, multiple choice questions and matching questions. The teacher or the students can edit the item bank by means of the editing interface. The parameter setting function enables the teacher to set the time period of a round and the challenge level of the competition mode. For example, the teacher can set the period of a round to three minutes, or the next item will be shown when the three students answer correctly.

Individual competition mode

Competition is an activity in which the students compete with their peers over a set period of time. To further provide different individual competition activities, the individual competition model was further split into an individual task competition (ITC) activity and an individual rush competition (IRC) activity.

ITC activity

In the ITC activity, (refer to Figure 3-a), the shared screen is divided into several zones equivalent to the number of participants. The students can do the multiple choice question exercises individually with their own cursors in their personal zone areas at their own speed. The whole exercise ends only when all students are finished. A personal progress bar to indicate individual advancement is displayed in their personal zone areas. Each student in the group can be aware of their peers’ progress on the shared display.

IRC activity

In the IRC activity, (refer to Figure 3-b), the multiple choice questions are displayed in the central area for all students, but the students answer with their own mice in their personal areas. The IRC activity can be set as a one-chance sub-activity, a two-chance sub-activity or a three-chance sub-activity. Only the fastest students can be awarded points during each round. For example, in the one-chance sub-activity, only one student can get a point. As soon as he/she gets the point, the question changes immediately.

Figure 3. Individual competition mode
Comparing the ITC and IRC activities, the competition intensity of IRC is higher than ITC because the students have to compete with others to choose the right answers within seconds. Among the IRC three levels, one-chance sub-activity, two-chance sub-activity and three-chance sub-activity, the three-chance sub-activity has the lowest intensity. The one-chance mode is set as the highest intensity which only allows a single person to give the right answer. The students should not only figure out the answer but also react faster than others.

**Group competition mode**

A quality classroom environment should provide several opportunities for social and group interaction. For group interaction, the MUSCLE system provides three group competition activities. They are Group Coop (GCO) activity, Group Competition (GCM) activity, and Group Rush Competition (GRC) activity.

**GCO activity**

GCO represents Group Coop activity which is a cooperative mode. All students are regarded as a large group. They cooperate to complete tasks proposed by the computer. In this mode, refer to Figure 4 (a), all the students’ cursors with their names are displayed in the same zone without boundaries on the shared display. The students are required to pick the Chinese words positioned in the center of the screen and drag the Chinese words to the appropriate English words. Within the allotted time, the set of questions will change if the students can make proper matches. No rushed time or rule constraints are applied in this mode.

**GCM activity**

GCM represents Group Competition activity. In this mode, refer to Figure 4 (b), the students are divided into two teams to partake in the activity, but all the students’ cursors with their names are displayed on the shared display in the same zone. The members of the two groups are required to recognize the Chinese words present on the shared display and drag the Chinese words to the appropriate English words. The team will be given one point if one of their team members drags the Chinese word to the appropriate English word. However, the proposed Chinese words are limited and will change if the students from two teams can match most of the questions. The team that has the highest score at the end of the game is the winner.

**GRC activity**

GRC represents Group Rush Competition activity. Refer to Figure 4 (c), the display is divided into two zones. Similar to the GCM mode, the students are also grouped into two teams. In GRC activity two groups of students are assigned to their group zones and the students cannot cross the boundary. In their own zone, the students can finish the exercise with their group members. The team members drag the Chinese words to the appropriate English words in their team space. The difference between GCM and GRC modes is that each team has their own questions in their team space and only the group members can answer their own questions. No matter how many points each team member earns, the winner is the team that has the highest end score.

![GCO: Group Coop, GCM: Group Competition, GRC: Group Rush Competition](image)

*Figure 4. Group competition mode*

To summarize, from the perspective of competition intensity, the GRC activity is higher than the GCM activity, and the GCM activity is higher than the GCO activity. From the group competition goal perspective, the GRC has the most clear competition goal and the GCO has the most loose competition goal.
Evaluation

To study the students’ preferences toward the different multiple-mice technology supported competition activities, sixty 7th grade students from two classes were involved in the study. Among them, thirty are female, and thirty are male. The subject is English vocabulary. Three 7th grade English units of approximately seventy words were involved. The three lessons changed as the English course proceeded. One lesson was old; another was new; the other was untaught.

A measurement of students’ learning achievement was involved in this research since a student’s level of achievement will affect his/her competition performance. According to the aim of the study, the competition activities were grouped by individual competitions and group competitions, and the students were cataloged as high-achieving students, average-achieving students and low-achieving students. The monthly English test results of the students were collected as the index of the students’ learning achievement. The grades of high-achieving students are the top one-third of the class and the low-achieving students are the bottom one-third of the class. The average-achieving students are in the middle.

Evaluation framework and process

In order to study the students’ competitive preferences among these different competition activities, an evaluation framework was applied. The evaluation framework was composed of three studies derived from a quantitative questionnaire shown as Figure 5. The first study focused on the students’ preferences among the individual competition activities and group competition activities; the second study measured the students’ preferences toward individual competition activities; the third study looked at the students’ preferences toward group competition activities. In the process of the evaluation, the questionnaire was applied to assess students’ behavior and competition preferences.

![Figure 5. Evaluation framework](image)

The perception questionnaire was designed to evaluate the students’ preferences. The questions cover subject preference, system usability, and competition activity preference factors. The evaluation was designed as three phases. Before the phases began, there was a warm-up session. This study was aimed at finding information concerning students’ competition preferences.

All of the students were involved in the activities. In the first week, the students were allowed to play around with the MUSCLE system as a warm up and began to understand how the activities would proceed. Each phase was enacted for two weeks to allow the students to become well-accustomed to the five different activities. Six students, as a group, perform the activities on the same screen at the same time. This means every student has a chance to practice each of the different competition modes at least twice.

Seven rounds of using the system were carried out. The first round was given to permit the students to warm up and allow the students to become familiar with the system. Then rounds two and three executed the first phase of the study. Each phase had two rounds and used the same mode twice. The scope of each phase was the same, but the two rounds were different in the exchange of two languages.
### Table 2. Practice schedule

<table>
<thead>
<tr>
<th>Round</th>
<th>System Activity</th>
<th>Unit</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Warm up 10 minutes</td>
<td>U4</td>
<td>Where’s your dog?</td>
</tr>
<tr>
<td>2</td>
<td>ITC, IRC (3/6)</td>
<td>U5</td>
<td>Are there any koalas in the zoo?</td>
</tr>
<tr>
<td>3</td>
<td>ITC, IRC (3/6)</td>
<td>U6</td>
<td>He’s dancing with his grandma.</td>
</tr>
<tr>
<td>4</td>
<td>IRC (1/6), IRC (2/6), IRC (3/6)</td>
<td>U5</td>
<td>Are there any koalas in the zoo?</td>
</tr>
<tr>
<td>5</td>
<td>IRC (1/6), IRC (2/6), IRC (3/6)</td>
<td>U6</td>
<td>He’s dancing with his grandma.</td>
</tr>
<tr>
<td>6</td>
<td>GCO, GCM, GRC</td>
<td>U7</td>
<td>What time is it?</td>
</tr>
<tr>
<td>7</td>
<td>GCO, GCM, GRC</td>
<td>U7</td>
<td>He’s dancing with his grandma.</td>
</tr>
</tbody>
</table>

### Evaluation results

In order to have a clear idea of what the students’ competition preferences are, the students were asked to give a ranking of the seven sub-activities according to their preference after filling in the five-point perception Likert item questionnaire. The students ranked each of the seven activities from one to seven. An activity with a rank of seven indicated the activity the student liked the most. If the student ranked the activity with a one, he/she did not prefer it at all. This ranking system shows what their preferences are. The students’ competition preferences toward different activities are elaborated below.

### Study I: The preference of the students toward individual and group competition mode

The first study was proposed to investigate the preferences of the high-, average- and low-achieving students toward the individual and group competition modes in general. As presented in Table 3, only the low-achieving students showed a significant difference in preference between individual and group competition modes (t(19) = 2.56, p < .05). The results indicate that the low-achieving students liked to perform with group members to work out the answers rather than answer the questions individually.

<table>
<thead>
<tr>
<th>Table 3. Students’ preference toward individual and group competition modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Competition Mode</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>High-Achieving</td>
</tr>
<tr>
<td>Average-Achieving</td>
</tr>
<tr>
<td>Low-Achieving</td>
</tr>
</tbody>
</table>

* p<.05

### Study II: The preference of the students toward individual competition sub-activities

The goal of the second study is to investigate the students’ preferences toward different individual competition sub-activities in which the ITC activity and IRC activity were involved. In the ITC, the students were asked to answer the questions at their own pace. On the contrary, in the IRC mode the students competed with others to strive for the chance to give answers. The competition levels in the IRC mode can be set as a one-chance sub-activity, a two-chance sub-activity and a three-chance sub-activity. Among the three levels, the three-chance sub-activity has the lowest intensity in which students had to compete with others to be one of the first three people who gave the right answer. The one-chance sub-activity was set as the highest intensity which only allowed a single person to give the right answer.

Based on the questionnaire, there was a significant difference in preference toward the ITC activity (F(2, 59) = 4.52, p < .015) between the three groups but no other significant differences were found in the three competition sub-activities of the IRC activity (see Table 4). The results showed that the high-achieving group preferred the ITC
activity most (mean = 6.15) compared to the low-achieving group (mean = 4.25), indicating that the high-achieving students liked to answer the questions at their own pace.

Table 4. Students’ preference toward individual competition sub-activities across levels

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITC Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>36.100</td>
<td>2</td>
<td>18.050</td>
<td>4.522*</td>
<td>High-Achieving (mean n = 6.15) &gt; Low-Achieving (mean = 4.25)</td>
</tr>
<tr>
<td>Within Groups</td>
<td>227.500</td>
<td>57</td>
<td>3.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>263.600</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRC Three-Chance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>3.033</td>
<td>2</td>
<td>1.517</td>
<td>.618</td>
<td>No Significant Difference</td>
</tr>
<tr>
<td>Within Groups</td>
<td>139.950</td>
<td>57</td>
<td>2.455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>142.983</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRC Two-Chance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.400</td>
<td>2</td>
<td>.200</td>
<td>.064</td>
<td>No Significant Difference</td>
</tr>
<tr>
<td>Within Groups</td>
<td>178.200</td>
<td>57</td>
<td>3.126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>178.600</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRC One-Chance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>5.833</td>
<td>2</td>
<td>2.917</td>
<td>.742</td>
<td>No Significant Difference</td>
</tr>
<tr>
<td>Within Groups</td>
<td>224.100</td>
<td>57</td>
<td>3.932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>229.933</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Table 5 presents the results concerning the attitudes toward individual competition sub-activities for each group. When closely examining the attitudes toward individual competition activities for each group, the high-, average- and low-achieving students preferred the ITC activity to the IRC activity, indicating that most of the students liked to answer the questions at their own pace instead of striving for a single chance to answer. Between the three competition intensity levels of the IRC activity, the results seemed to show that high-, average- and low-achieving students prefer a moderately competitive mode rather than an intensely competitive mode.

Table 5. Students’ preference toward individual sub-activities for each level

<table>
<thead>
<tr>
<th></th>
<th>ITC</th>
<th>IRC One-Chance</th>
<th>IRC Two-Chance</th>
<th>IRC Three-Chance</th>
<th>Paired t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>SD</td>
<td>SD</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>SD</td>
<td>SD</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>High-Achieving</td>
<td>6.15</td>
<td>3.05</td>
<td>3.40</td>
<td>3.30</td>
<td>3.181**</td>
</tr>
<tr>
<td>SD</td>
<td>1.87</td>
<td>1.85</td>
<td>1.67</td>
<td>1.92</td>
<td></td>
</tr>
<tr>
<td>Average-Achieving</td>
<td>5.20</td>
<td>2.55</td>
<td>3.20</td>
<td>3.85</td>
<td>4.535**</td>
</tr>
<tr>
<td>SD</td>
<td>1.94</td>
<td>2.33</td>
<td>1.82</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>Low-Achieving</td>
<td>4.25</td>
<td>2.38</td>
<td>3.30</td>
<td>3.80</td>
<td>2.156*</td>
</tr>
<tr>
<td>SD</td>
<td>2.17</td>
<td>1.72</td>
<td>1.81</td>
<td>1.36</td>
<td></td>
</tr>
</tbody>
</table>

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

Study III: The preference of the students toward group competition activities

The goal of this study is to investigate the students’ preferences toward group competition activities. Three group competition activities, GCO, GCM and GRC, were involved in this study. Among the three activities, GCO is the lowest intensity competition activity, and GRC is the highest intensity competition activity. The results showed that the three achieving groups showed no significant differences in preference toward the three group competition activities (see Table 6).
Table 6. Students’ preference toward group competition activities across levels

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GCO Activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>8.633</td>
<td>2</td>
<td>4.317</td>
<td>1.618</td>
<td>No Significant</td>
</tr>
<tr>
<td>Within Groups</td>
<td>152.100</td>
<td>57</td>
<td>2.668</td>
<td></td>
<td>Difference</td>
</tr>
<tr>
<td>Total</td>
<td>160.733</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GCM Activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>14.433</td>
<td>2</td>
<td>7.217</td>
<td>2.770</td>
<td>No Significant</td>
</tr>
<tr>
<td>Within Groups</td>
<td>148.500</td>
<td>57</td>
<td>2.605</td>
<td></td>
<td>Difference</td>
</tr>
<tr>
<td>Total</td>
<td>162.933</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GRC Activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>16.933</td>
<td>2</td>
<td>8.467</td>
<td>1.748</td>
<td>No Significant</td>
</tr>
<tr>
<td>Within Groups</td>
<td>276.050</td>
<td>57</td>
<td>4.843</td>
<td></td>
<td>Difference</td>
</tr>
<tr>
<td>Total</td>
<td>292.983</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantitative questionnaire study

A five-point perception Likert item questionnaire shown as Table 7 was applied to study the sixty students’ English learning attitudes and competition preferences. One or two students missed some items in filling out the questionnaire that makes the summary in some items are not exactly equal to one hundred percent. According to the statistics, the students expressed that they liked the English course (Item 1), but only 2.75 students said that they would like to memorize English vocabularies (Item 2). English vocabularies are important for students in learning English, but the willingness of the students to memorize English vocabularies in traditional way was not so high. The questionnaire results also revealed that almost half of the students expressed that their English achievement was not good (Item 3). However, the students showed that they like to learn words with the MUSCLE system (Item 4), and the students expressed that they can learn words with the MUSCLE system (Item 5). The students also expressed that they would like to practice the exercises with the MUSCLE system (Item 7). It is interesting that the students expressed that they don’t like to memorize English vocabularies in traditional way (Item 2), but they can accept to learn words in MUSCLE system (Item 4 & Item 5). In general, the students can accept the MUSCLE system (Item 6 & 7), and they can also accept applying the competition mechanism (Item 8 & 9) in the classroom.

With MUSCLE system, the motivation to learn vocabulary was promoted, and the students had fun in participating in MUSCLE learning activities (Item 4 & 6). Furthermore, most of them are willing to learn with others by using MUSCLE. Besides, more than half of the students like the competition mechanism (Item 8 & 9) though there are some various preferences. The questionnaire results roughly indicated that the students need support in learning English, and they could accept the MUSCLE system as a kind of learning support system.

Table 7. Statistics of the students’ preference for MUSCLE (n=60)

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly agree (%)</th>
<th>Agree (%)</th>
<th>Neutral (%)</th>
<th>Disagree (%)</th>
<th>Strongly disagree (%)</th>
<th>Ave.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I like to have English course.</td>
<td>27.9</td>
<td>27.9</td>
<td>31.1</td>
<td>8.2</td>
<td>4.9</td>
<td>3.66</td>
<td>1.12</td>
</tr>
<tr>
<td>2. I like to memorize English vocabularies.</td>
<td>8.2</td>
<td>18.0</td>
<td>34.4</td>
<td>16.4</td>
<td>21.3</td>
<td>2.75</td>
<td>1.23</td>
</tr>
<tr>
<td>3. My English achievement is great.</td>
<td>11.5</td>
<td>11.5</td>
<td>27.9</td>
<td>27.9</td>
<td>21.3</td>
<td>2.64</td>
<td>1.27</td>
</tr>
<tr>
<td>4. I like to learn words with MUSCLE system.</td>
<td>39.3</td>
<td>23.0</td>
<td>32.8</td>
<td>1.6</td>
<td>3.3</td>
<td>3.93</td>
<td>1.05</td>
</tr>
<tr>
<td>5. I can learn new words from the activities.</td>
<td>33.0</td>
<td>20.0</td>
<td>30.0</td>
<td>6.7</td>
<td>10.0</td>
<td>3.6</td>
<td>1.29</td>
</tr>
<tr>
<td>6. It is fun that six people do the practice altogether.</td>
<td>42.6</td>
<td>13.1</td>
<td>31.1</td>
<td>6.6</td>
<td>4.9</td>
<td>3.83</td>
<td>1.21</td>
</tr>
<tr>
<td>7. If I have chances, I’d like to practice with MUSCLE system.</td>
<td>36.7</td>
<td>23.3</td>
<td>30.0</td>
<td>5.0</td>
<td>5.0</td>
<td>3.82</td>
<td>1.14</td>
</tr>
<tr>
<td>8. Competition makes the practice more interesting.</td>
<td>11.5</td>
<td>18.0</td>
<td>36.1</td>
<td>21.3</td>
<td>11.5</td>
<td>3.34</td>
<td>1.33</td>
</tr>
<tr>
<td>9. I like to compete with my own group members.</td>
<td>13.1</td>
<td>21.3</td>
<td>42.6</td>
<td>14.8</td>
<td>8.2</td>
<td>3.16</td>
<td>1.10</td>
</tr>
</tbody>
</table>
Discussions

Different competition modes of MUSCLE have been applied in a classroom, and the students’ preferences toward different competition modes have been explored. In this section, the discussions of technology accessibility in classroom, MUSCLE system environment for teachers and students, and different levels of students’ learning achievements preferences toward different competition modes are elaborated below.

MUSCLE provides affordable information technology accessibility in classrooms

System usability is the critical linchpin in a technology-enhanced classroom. Teachers and students cannot tolerate unstable and low usability system in the classroom. MUSCLE has two versions. The previous version was designed using wireless sensor technology (Chang & Chen, 2010), and the later version was implemented through multiple-mice technology. Ideally, the MUSCLE system can connect more than one hundred mice to a computer simultaneously. However, in real practice, the MUSCLE system works best when less than ten students use mice with a computer at the same time. Using less than ten mice, the experimental participants express a high usability of the MUSCLE system. The MUSCLE system provides an affordable new opportunity for teachers and students to interact in the classroom with minimal information technology accessibility.

MUSCLE system usability is acceptable to teachers and students

According to the questionnaire results, the students showed a positive attitude toward the MUSCLE system. Most of the students liked to learn vocabulary with classmates by using MUSCLE and they felt interested in using multiple mice together, indicating the MUSCLE system can be accepted by the students in classroom. However, few students expressed an opposite opinion about this system, which reveals a concern that the equipment used in the MUSCLE system was affordable, but it may not meet all of the students’ needs for learning. Students may move more freely and engage in activities using their whole body if MUSCLE is supported by more advanced technology such as wireless mice, big screens and high-performance CPUs. The motivation of learning may be promoted and the level of learning achievement may be improved. Besides, the teacher’s observation indicated that the students would ask about MUSCLE system and show their willingness to use it. Many students surrounded the screen and the students fully engaged themselves. Some of them would look up new words in the textbook to answer more correctly. In specific, the students who generally lacked interest in English would help the teacher to pack up the equipment after classroom and talk about the activities.

High-achieving and low-achieving students’ preferences for competition modes are different

According to the studies, the low-achieving students showed a preference for group competition modes and high-achieving students preferred the ITC activity the most. In class, the circumstances that students have different preferences and different achievements does make a difference. It has a crucial effect on teachers’ instructional design. The results of this study may remind teachers of students’ different preferences for competition. How to design a suitable lesson design for the whole class is a big issue for all teachers. In the classroom, the class cannot be divided into two or more parts. Nevertheless, the needs of different students should be considered by all teachers. However, it is not easy for teachers to individually find a balance. Fortunately, technology may provide a large amount of help for teachers. Teachers who can make good use of instructional technology, such as the MUSCLE system, may greatly aid students. Integrating suitable technology with instruction, students’ level of achievement may become better and better.

Conclusions

Competition is an instinctive human behavior and takes a role in our daily life. Excessively competitive activities can have many negative effects and no doubt frustrate many students. However, appropriate competition intensity can improve the quality of classroom interaction. A competitive activity is generally considered an effective technique in motivating people to learn and to excel (Yu, Chang, Liu & Chan, 2002). Adding an element of competition is widely
believed to be a motivation-enriching strategy in play, work, and education (Deci et al., 1981). Social learning researchers have indicated the importance of applying interactive activities in classrooms and it has been shown that interaction between students can be enhanced through well-design competitive activities in a technology-integrated learning environment. However, the students’ ability to readily access technology was formerly an obstacle in a technology-enhanced classroom. This study demonstrated the potentials of using MUSCLE to motivate students to memorize English vocabulary. Using MUSCLE, each student can have a mouse in hand to interact with their peers. Instead of equipping the classrooms with more computers, MUSCLE allows students to interact with their peers and compete with other groups on a shared display.

In this study, combining multiple-mice technology and competition activities design, the MUSCLE system was designed, implemented and applied in a classroom. In addition, an evaluation framework was practiced in this study. The framework covered students’ learning achievement, individual competition, and group competition. Three study designs were applied based on the framework, which covered the preferences of the students toward individual competition activities, the preferences of the students toward group competition activities, the preferences of the students toward individual and group competition through a quantitative questionnaire. With the system, sixty 7th grade students participated in this study through analyzing English vocabulary. The results showed that the students’ competition preferences were related to their level of achievement. Comparing to the individual competition mode and group competition mode, the low-achieving students preferred group competition mode significantly over individual competition mode. It seems to indicate that low-achieving students need more scaffolding provided by more-capable peers when accomplishing tasks. In addition, among different individual competitive activities, the low-achieving students preferred the ITC activity which is a less competitive activity. The results indicated that the three achieving groups showed no significant differences in preference toward the three group competition activities.

Before the experiment, the five-point perception Likert item questionnaire results indicated that the students expressed that their English achievement was not good, and most of them didn’t like to memorize English vocabulary in traditional way. After the experiment, the students expressed that they could accept practicing the exercises with the MUSCLE system in the classroom and compete with their peers on a shared notebook with their own mouse simultaneously.

The MUSCLE system demonstrated the potential of using multiple-mice technology in a classroom, and the system evaluation results indicated that the MUSCLE system can provide affordable information technology accessibility in a classroom. By using the multiple-mice technology, all students can have access to information technology at minimal cost to the teachers. The contribution of this study is to demonstrate the potential of applying multiple-mice technology into a regular classroom and providing five competition activities for the teachers and the students to apply. Through this study, the authors found that different learning achievement students have different preferences in the multiple-mice supported classroom, and it is interesting that the high-achieving students have an opposite preference to the low-achieving students. As instructors or designers, more trade-offs should be considered. This was pioneer study with sixty 7th grade students in a rural area in Taiwan where culture differences and urban-rural gap issues may come into play. This is also a study demonstrating the possibility of using non-PC-like human-interaction technology in a real teaching field. More novel technologies such as gestures-based computing, wireless sensors, multiple-touch technology would be available in the future to help technology-enhanced classroom learning. In the next study, the authors are trying to exploring the possible classroom solutions to compromise on these differences.

**Acknowledgements**

This project was supported in part by the National Science Council of Taiwan, Republic of China under the grant NSC 99-2511-S-415-006-MY2. The authors particularly thank Chien Wen Chen for his enthusiastic participation in the system implementation.

**References**


Developing a New Computer Game Attitude Scale for Taiwanese Early Adolescents

Eric Zhi-Feng Liu, Chun-Yi Lee* and Jen-Huang Chen

Graduate Institute of Learning and Instruction, Center of Teacher Education, Research Center for Science and Technology for Learning, National Central University, Jhongli City, Taoyuan, Taiwan // Tzu Chiang Junior High School, Jhongli City, Taoyuan, Taiwan // Ruei Fong Elementary School, Hsinchu County, Taiwan // toten@cc.ncu.edu.tw // chunyi.lii@gmail.com // hs0888@mail.nc.hcc.edu.tw

*Corresponding author

(Submitted April 05, 2011; Revised August 27, 2011; October 10, 2011)

ABSTRACT

With ever increasing exposure to computer games, gaining an understanding of the attitudes held by young adolescents toward such activities is crucial; however, few studies have provided scales with which to accomplish this. This study revisited the Computer Game Attitude Scale developed by Chappell and Taylor in 1997, reworking the overall structure of the instrument, and increasing the number of items to 22. The revised scale covers five factors: learning, confidence, liking, participation, and leisure, grouped into three subscales of cognition, affection and behavior. Data gathered from 354 elementary school students has demonstrated the validity and reliability of the proposed scale for investigating the attitudes of elementary students toward learning through computer games, providing additional insight into the influence of gender and Internet usage.

Keywords

Computer game, Computer game attitude scale, Taiwanese early adolescents, Early adolescents

Introduction

Playing games is a natural activity for elementary school students (or early adolescents), helping them to understand the world and widen their experience by exploring the full range of possibilities open to them (Eisner, 1982). The social skills of children are developed through interaction with their playmates, and their desires and needs are fulfilled through the games they engage in, enabling them to develop into healthy well-adjusted adults (Montessori, 1917). According to educational psychologist Jean Piaget, human cognitive development from birth to age fourteen, is divided into four stages: the sensorimotor stage, preoperational stage, concrete operational stage, and formal operational stage (Dembo, 1994). Sixth graders (approximately 13 years old), are in the formal operational stage of development, fully capable of the abstract thought required to cognize abstractions represented in a virtual computer environment. In recent years, computer games have become one of the most popular leisure activities for adolescents and young adults, providing a space in which to develop critical thinking and engage in active learning (Griffiths, 2010). Virtual worlds are perceived as a safe environment, in which learners can develop skills, concepts, and strategies. The limitless potential of computer games in education has been widely recognized and the notion of developing games with more serious themes has emerged in the fields of education and training (Prensky, 2001). Many researchers have attempted to incorporate computer games into school curricula as a means to enhance academic performance (Lee & Chen, 2009), cognitive development (Jenkins, 2002), motivation toward learning (Liu & Lin, 2009), and attention span and concentration (McFarlane, Sparrowhawk, & Heald, 2002). Computer games enhance the process of learning by promoting the use of visualization, experimentation, and creativity in play (Betz, 1995). Computer games also challenge students and provide them with feedback, while allowing them to satisfy their curiosity and earn a sense of achievement (Chiang, Lin, Cheng, & Liu, 2011; Liu & Lin, 2009). Imaginary worlds provide an environment in which students may fulfill their basic human needs, experiencing the taste of victory, the joy of success, and the satisfaction that comes from collaboration or competition with others. Satisfying the instinctual yearnings of participants makes learning more interesting and imbues the process with meaning (Cagiltay, 2007; Gee, 2003; Liu & Lin, 2009).

Based on the viewpoint of learning theory, behaviorist Doob (1947) considered attitude an implicit response, which is both anticipatory and mediating in reference to patterns of overt responses. In this manner, attitudinal responses do not differ from other responses and can be gained through conditioning. Social psychologist Allport (1935) defined attitude as a mental and neural state of readiness, exerting a directive or dynamic influence on the response of individuals to all objects and situations within which they are situated. For instance, the mental readiness of a runner...
anticipating the sound of the starting pistol is an attitude. Thurstone and Chave (1928) defined attitude as a scale of affection related to one’s opinions, in which particular objectives are favored or rejected. In principle, attitudes are personal; however, they occasionally fulfill a social function. People growing up in the same culture and society are influenced by the same cultural traditions, social norms, customary practices, and other factors. It is not surprising that in such circumstances individuals tend to exhibit similar traits and a high degree of consistency in their attitudes. Attitudes are learned, and as such, are closely related to one’s experiences in the process of learning. In summary, attitude can be defined as the outward manifestation of an individual’s evaluation of an entity, based on previous knowledge and beliefs.

In educational research, the attitude of students towards subject matter is often viewed as an important variable in predicting performance. It would therefore be unwise to suggest integrating computer games into classrooms without fully investigating how students view them (Bourgonjon, Valcke, Soetaert, & Schellens, 2010). Reece and Gable (1982) proposed that computer attitude comprises three components: cognition, affection and behavior. Brown, Brown, and Baack (1988) also described how computer attitude includes a behavioral component, entailing the specific actions of an individual toward computers; an affective component, representing the internal feeling of an individual about computers; and a cognitive component, comprising the beliefs of an individual regarding computers. Campbell (1987) pointed out that computer attitude is the general impression, opinions, and cognition of computers. A summary of previous research results has led to the conclusion that computer attitude comprises cognitive, affective, and behavioral components. The cognitive component represents an individual’s impression and opinion of computers. The affective component represents an individual’s personal feelings about computers. The behavioral component represents specific behaviors, such as an inclination to use computers.

Although a number of instruments have been developed to measure attitudes toward computers, far fewer instruments have been developed for measuring attitudes toward computer games. Bonanno and Koomers (2008) developed a tool to measure four components of attitude toward gaming: affective components, perceived control, perceived usefulness, and behavioral components. Their survey, including 21 statements scored on a 5-point Likert scale, was used with a sample of college students to investigate the influence of gender and gaming competence on attitudes toward gaming. However, this instrument requires further refinement and validation to ensure reliability and construct validity. Chappell and Taylor (1997) developed an instrument called the Computer Game Attitude Scale (CGAS) to evaluate the attitudes of students toward educational computer games. Their study provided evidence supporting the reliability and factorial validity of the scores of the CGAS and its two subscales: comfort and liking. However, using principle component factor analysis indicated that both the comfort and liking subscales explained only 44% of all variables. At less than 50%, this scale clearly requires further validation.

Previous studies (Bonanno & Koomers, 2008; Chappell & Taylor, 1997) addressing attitudes toward computer games have tended to emphasize the affective component. However, the integration of information technology and computer games provides early adolescents with another avenue for learning (Jonassen, 1996). There are a great many types of computer games, but not all of them have a practical application in education (Can & Cagiltay, 2006). For this reason, game playing as well as game based learning should be considered within the context of schooling. Developing an instrument for measuring attitudes toward computer based activities would enable the expansion of the discussion beyond the aspect of affection, to include cognition and behavior regarding computer games (Adcock & Van Eck, 2005). This study built upon the model by Reece and Gable (1982), to develop a New Computer Game Attitude Scale (NCGAS) for elementary school students, by revising the CGAS and adding new items based on the following three subscales: (1) a cognition subscale representing the impressions and opinions of individuals regarding the use of computer games; (2) an affection subscale representing the feelings of individuals regarding computer games; and (3) a behavior subscale representing specific behaviors, such as an inclination to use computers.

Previous studies have suggested that separating the experience students have with computer games from gender issues is nearly impossible. Bonanno and Koomers (2008) described how males tend not only to play games more often, but to play different types of games. Males also tend to exhibit significantly different attitudes toward the use of computer games (Bonanno & Koomers, 2008). In contrast, Karakus, Inal, and Cagiltay (2008) found no significant gender differences in whether games are viewed as useful in education. Gender differences may already be changing, and with the ongoing expansion of online gaming and considerable increases in the number of female players, computer games may be equally effective and motivating for both male and female students (Papastergiou, 2009). Therefore, we also analyzed variables including gender and experience using the internet.
Method

Participants

The Computer Game Attitude Scale was developed for sixth and seventh graders; therefore, we selected 354 sixth graders as the participants in this study. The students were from five elementary schools in Taipei County, Hsinchu County, and Taichung City. Among the respondents, 168 were female and 166 male. Although the sample size of these young adolescents was relatively small, they were representative of various demographic and academic backgrounds.

Materials

To develop the New Computer Game Attitude Scale (NCGAS), items were adapted from Chappell and Taylor’s (1997) Computer Game Attitude Scale, with additional items contributed by the authors of this study. Chappell and Taylor (1997) proposed the following two subscales for computer game attitudes: comfort and liking, including a total of 20 items. The self-writing section listed 60 items, comprising cognition (including learning and confidence), affection (including liking), and behavior (including participation, leisure and negative behavior) toward computer games. All items were tested for face validity and content validity by an expert in the design of questionnaires and six highly experienced elementary school teachers. In the end, 60 items were included in the New Computer Game Attitude Scale. The 60 items were presented to 354 elementary students. Many elementary students select a neutral response without thinking; therefore, a four-point Likert scale in which 4 = strongly agree, 3 = agree, 2 = disagree, and 1 = strongly disagree was used to avoid this situation. Items demonstrating poor reliability or validity were deleted, leaving only 22 items in the New Computer Game Attitude Scale. The final questionnaire contained three subscales and five factors, and the cumulative explained variance of each exceeded 50%. Items were retained only when they exceeded +0.40 or were less than -0.40 for relevant factors and less than the absolute value of ±0.40 for non-relevant factors. The three subscales were cognition (learning and confidence), affection (liking), and behavior (participation and leisure). The five factors were learning (coded as LRN), confidence (CON), liking (LIKE), participation (PAR) and leisure (LEI) (refer to Table 1). A detailed description of the five factors is presented below:

1. Learning: used to measure the perception of students as per the positive impact of the computer game on learning.
2. Confidence: used to measure the confidence of students to independently control the computer game.
3. Liking: used to measure how much the students liked using the computer game.
4. Participation: used to measure the degree to which students actual participated in the activities related to the computer game.
5. Leisure: used to measure the degree to which the computer game was incorporated in the leisure time of the students and their perceptions of the game as a leisure activity.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Factor</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition</td>
<td>LRN</td>
<td>1. I was able to install other software on my computer after installing computer games.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. I learned how to look up software instructions by playing computer games.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. I would be more willing to take classes if computer games were used in class.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Playing computer games lets me use my imagination.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Playing computer games increases my typing speed.</td>
</tr>
<tr>
<td></td>
<td>CON</td>
<td>1. Playing computer games is not my specialty (-).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Playing computer games is easy to me.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. I am not good at playing computer games (-).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. I am skilled at playing computer games.</td>
</tr>
<tr>
<td>Affection</td>
<td>LIKE</td>
<td>1. I am very interested in solving problems in computer games.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. If there are any unresolved issues in computer games, I will continue thinking of them at another time.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. If I encounter a problem that I do not understand while playing computer games, I will keep trying until I find the answer.</td>
</tr>
</tbody>
</table>
Behavior PAR 1. If there is a computer game club in school, I will participate in it.
2. If there is a computer game camp at school, I will participate in it.
3. At a computer show, I would visit a computer game booth.

LEI 1. Playing computer games makes me feel happy.
2. Playing computer games is part of my life.
3. I want to play computer games after each examination.
4. I kill time by playing computer games.
5. I talk about computer games with my friends in my spare time.
6. I can chat with others when playing computer games and won’t feel lonely.

(-): The score of these items should be reversed when performing analysis

Results

General findings

The main aim of the present paper was to assess the construct validity of the attitudinal scales and to determine their reliability. In this study, exploratory principal components factor analysis with varimax rotation was used for analysis of validity and Cronbach’s alpha was used as for the analysis of reliability. For the cognition subscale, the eigenvalues of the factor, Learning, (exceeding 1: 2.651) and the factor, Confidence, (exceeding 1: 2.385) were determined using principal components analysis with varimax rotation, as shown in Table 2. These two factors accounted for 50.352% of the variance. The internal reliability index (alpha coefficient) was sufficient for Learning (0.737) and Confidence (0.77) as well as the entire subscale (0.819).

Table 2. Rotated factor loadings and Cronbach’s alpha values for the cognition subscale of the Computer Game Attitude Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1: Learning</th>
<th>Factor 2: Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRN1</td>
<td>0.537</td>
<td></td>
</tr>
<tr>
<td>LRN2</td>
<td>0.629</td>
<td></td>
</tr>
<tr>
<td>LRN3</td>
<td>0.557</td>
<td></td>
</tr>
<tr>
<td>LRN4</td>
<td>0.657</td>
<td></td>
</tr>
<tr>
<td>LRN5</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>LRN6</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Factor 1: Learning alpha = 0.737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON1</td>
<td></td>
<td>0.666</td>
</tr>
<tr>
<td>CON2</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td>CON3</td>
<td></td>
<td>0.813</td>
</tr>
<tr>
<td>CON4</td>
<td></td>
<td>0.715</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>2.651</td>
<td>2.385</td>
</tr>
<tr>
<td>% of variance</td>
<td>26.505%</td>
<td>23.847%</td>
</tr>
<tr>
<td>Overall alpha=0.819, total variance explained is 50.352%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the affection subscale, the eigenvalue of the factor Liking (exceeding 1: 1.964) was determined using principal components analysis without rotation, as shown in Table 3. The factor of Liking accounted for 65.455% of the variance. The internal reliability index (alpha coefficient) was sufficient for the factor of Liking (0.736).

Table 3. Factor loading and Cronbach’s alpha for the affection subscale of the Computer Game Attitude Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1: Liking</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIKE1</td>
<td>0.791</td>
</tr>
<tr>
<td>LIKE2</td>
<td>0.807</td>
</tr>
<tr>
<td>LIKE3</td>
<td>0.829</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>1.964</td>
</tr>
<tr>
<td>% of variance</td>
<td>65.455%</td>
</tr>
</tbody>
</table>
For the behavior subscale, the eigenvalues of the factor, Participation, (exceeding 1: 3.259) and the factor, Leisure, (exceeding 1.878) were determined using principal components analysis with varimax rotation, as shown in Table 4. These two factors accounted for 50.073 % of the variance. The internal reliability index (alpha coefficient) was sufficient for Participation (0.649) and Leisure (0.839), as well as the entire subscale (0.822).

Table 4. Rotated factor loadings and Cronbach’s alpha for the behavior subscale of the Computer Game Attitude Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1: Participation</th>
<th>Factor 2: Leisure</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR1</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>PAR2</td>
<td>0.751</td>
<td></td>
</tr>
<tr>
<td>PAR3</td>
<td>0.646</td>
<td></td>
</tr>
</tbody>
</table>

Factor 1: Confidence alpha = 0.839

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1: Participation</th>
<th>Factor 2: Leisure</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEI1</td>
<td>0.788</td>
<td></td>
</tr>
<tr>
<td>LEI2</td>
<td>0.734</td>
<td></td>
</tr>
<tr>
<td>LEI3</td>
<td>0.729</td>
<td></td>
</tr>
<tr>
<td>LEI4</td>
<td>0.729</td>
<td></td>
</tr>
<tr>
<td>LEI5</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>LEI6</td>
<td>0.731</td>
<td></td>
</tr>
</tbody>
</table>

Eigenvalue 3.259 1.878

% of variance 36.21% 20.863%

Overall alpha = 0.822, total variance explained is 50.073%

The intercorrelation matrix showed significant correlation coefficients among five factors of computer game attitude and three subscales (Table 5). The correlation coefficients among the five factors ranged between 0.405** and 0.669**, and all correlations reached a significance level of 0.01. The correlation coefficients among the three subscales ranged between 0.644** and 0.762**, and all of the correlations reached a significance level of 0.01. The three subscales and five factors were used to provide a coherent measurement of computer game attitude.

Table 5. Intercorrelation matrix of five computer game attitude factors and three subscales

<table>
<thead>
<tr>
<th>Factor</th>
<th>LRN</th>
<th>CON</th>
<th>LIKE</th>
<th>PAR</th>
<th>LEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRN</td>
<td>-</td>
<td>0.533**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CON</td>
<td>0.533**</td>
<td>-</td>
<td>0.482**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LIKE</td>
<td>0.645**</td>
<td>0.482**</td>
<td>-</td>
<td>0.562**</td>
<td>-</td>
</tr>
<tr>
<td>PAR</td>
<td>0.521**</td>
<td>0.434**</td>
<td>0.562**</td>
<td>-</td>
<td>0.405**</td>
</tr>
<tr>
<td>LEI</td>
<td>0.669**</td>
<td>0.555**</td>
<td>0.541**</td>
<td>0.405**</td>
<td>-</td>
</tr>
</tbody>
</table>

Subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Cognition</th>
<th>Affection</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>0.653**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Behavior</td>
<td>0.762**</td>
<td>0.644**</td>
<td>-</td>
</tr>
</tbody>
</table>

** P < 0.01

Table 6. Descriptive statistics of scores on the five factors and three subscales of the Computer Game Attitude Scale

<table>
<thead>
<tr>
<th>Factor</th>
<th>Items</th>
<th>Range</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRN</td>
<td>6</td>
<td>1-4</td>
<td>3.093</td>
<td>0.638</td>
<td>-0.803</td>
</tr>
<tr>
<td>CON</td>
<td>4</td>
<td>1-4</td>
<td>2.834</td>
<td>0.757</td>
<td>-0.231</td>
</tr>
<tr>
<td>LIKE</td>
<td>3</td>
<td>1-4</td>
<td>3.095</td>
<td>0.755</td>
<td>-0.783</td>
</tr>
<tr>
<td>PAR</td>
<td>3</td>
<td>1-4</td>
<td>2.528</td>
<td>0.797</td>
<td>0.049</td>
</tr>
<tr>
<td>LEI</td>
<td>6</td>
<td>1-4</td>
<td>3.105</td>
<td>0.706</td>
<td>-0.872</td>
</tr>
</tbody>
</table>

Subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Cognition</th>
<th>Affection</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition</td>
<td>10</td>
<td>1-4</td>
<td>2.99</td>
</tr>
<tr>
<td>Affection</td>
<td>3</td>
<td>1-4</td>
<td>3.095</td>
</tr>
<tr>
<td>Behavior</td>
<td>9</td>
<td>1-4</td>
<td>2.913</td>
</tr>
</tbody>
</table>
Student scores on the scale

Table 6 presents descriptive statistics or the five factors and the three subscales. Among the five factors, students scored highest on the Leisure factor, followed by the Liking and Learning, with Confidence and Participation scoring the lowest. Among the three subscales, students scored highest on the Affection subscale, followed by Cognition and Behavior. These results imply that these students tended to learn from the computer game, liked using the computer game, and played the computer game in their leisure time. The relatively lower scores for the Confidence and Participation factors suggest that some of the students might have had difficulty playing the computer game alone and joining related activities. For elementary school students, it is difficult to pay the fees required to join related activities or independently accomplish the difficult missions encountered while playing the computer game.

Gender differences on the scale

This study compared the scores of male and female students for the five factors and three subscales of the NCGAS. The results of t-tests are presented in Table 7, revealing that male students expressed stronger positive attitudes toward the computer game than their female counterparts for all five factors and three subscales. In other words, the male students believed more strongly that computer games have a positive impact on learning and showed greater confidence, liked the games more, and participated in the games more than the females students did. The male students were also likely to use the computer games more frequently than female students in their leisure time.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Gender</th>
<th>Mean</th>
<th>S. D.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRN</td>
<td>Male</td>
<td>3.028</td>
<td>0.722</td>
<td>3.966***</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.746</td>
<td>0.568</td>
<td></td>
</tr>
<tr>
<td>CON</td>
<td>Male</td>
<td>3.002</td>
<td>0.842</td>
<td>7.207***</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.349</td>
<td>0.811</td>
<td></td>
</tr>
<tr>
<td>LIKE</td>
<td>Male</td>
<td>3.147</td>
<td>0.717</td>
<td>5.833***</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.703</td>
<td>0.672</td>
<td></td>
</tr>
<tr>
<td>PAR</td>
<td>Male</td>
<td>2.909</td>
<td>0.881</td>
<td>4.301***</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.522</td>
<td>0.758</td>
<td></td>
</tr>
<tr>
<td>LEI</td>
<td>Male</td>
<td>3.278</td>
<td>0.689</td>
<td>5.132***</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.885</td>
<td>0.712</td>
<td></td>
</tr>
<tr>
<td>Subscale</td>
<td>Cognition</td>
<td>Male</td>
<td>3.017</td>
<td>0.714</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.587</td>
<td>0.583</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Affection</td>
<td>Male</td>
<td>3.147</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.703</td>
<td>0.673</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Behavior</td>
<td>Male</td>
<td>3.155</td>
<td>0.636</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>2.764</td>
<td>0.611</td>
<td></td>
</tr>
</tbody>
</table>

***p < 0.001

Internet experience and NCGAS

This study also examined the relationship between the attitudes of students toward computer games and their experience using the Internet. Internet experience is defined in this study as the weekly Internet usage by students at home. We divided the sample into five groups according to their Internet experience: no usage, less than two hours, 2–4 hours, 4–6 hours, and more than 6 hours. In the sample, 12.7 % of the students did not use the Internet at home, 23.2 % used the Internet for less than 2 hours per week, 32.2 % for 2–4 hours, 13.8 % for 4–6 hours, and 18.1 % for more than 6 hours. Table 8 compares the groups according to Internet experience and their attitudes toward the computer game. ANOVA testing revealed that Internet experience played a role in the factors of Learning, Confidence, Liking, Participation and Leisure, and influenced the subscales of Cognition, Affection, and Behavior. Through a series of Scheffe tests, it was determined that students who use the Internet more tended to have statistically higher scores on the Cognition, Affection and Behavior subscales as well as the factors of Learning, Confidence, Liking, Participation, and Leisure. These relationships indicate that the students with more Internet
experience tended to have positive perceptions of learning with computer games, demonstrate a high level of confidence playing them, enjoy these activities, and participate in them both during class as well as in their leisure time. Furthermore, they tend to have higher scores for Cognition, Affection, and Behavior related to learning with computer games. Increasing one’s experience with the Internet could greatly help students to develop favorable attitudes toward using computer games for learning.

### Table 8. Analysis of Internet experience and NCGAS

<table>
<thead>
<tr>
<th>Internet experiences</th>
<th>LRN</th>
<th>CON</th>
<th>LIKE</th>
<th>PAR</th>
<th>LEI</th>
<th>Cognition</th>
<th>Affection</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) no usage</td>
<td>2.304 (0.779)</td>
<td>2.069 (0.807)</td>
<td>2.592 (0.749)</td>
<td>2.533 (0.853)</td>
<td>2.492 (0.804)</td>
<td>2.210 (0.725)</td>
<td>2.592 (0.749)</td>
<td>2.506 (0.703)</td>
</tr>
<tr>
<td>(2) less than 2 hours</td>
<td>2.720 (0.679)</td>
<td>2.503 (0.973)</td>
<td>2.878 (0.765)</td>
<td>2.638 (0.775)</td>
<td>2.815 (0.754)</td>
<td>2.633 (0.714)</td>
<td>2.878 (0.765)</td>
<td>2.756 (0.688)</td>
</tr>
<tr>
<td>(3) 2–4 hours</td>
<td>2.915 (0.500)</td>
<td>2.568 (0.745)</td>
<td>2.877 (0.677)</td>
<td>2.655 (0.722)</td>
<td>3.091 (0.634)</td>
<td>2.776 (0.517)</td>
<td>2.877 (0.677)</td>
<td>2.945 (0.557)</td>
</tr>
<tr>
<td>(4) 4–6 hours</td>
<td>3.119 (0.538)</td>
<td>2.883 (0.731)</td>
<td>3.109 (0.636)</td>
<td>2.823 (0.948)</td>
<td>3.361 (0.538)</td>
<td>3.025 (0.542)</td>
<td>3.109 (0.636)</td>
<td>3.181 (0.546)</td>
</tr>
<tr>
<td>(5) more than 6 hours</td>
<td>3.263 (0.573)</td>
<td>3.434 (0.618)</td>
<td>3.214 (0.731)</td>
<td>3.026 (0.964)</td>
<td>3.625 (0.447)</td>
<td>3.331 (0.523)</td>
<td>3.214 (0.731)</td>
<td>3.425 (0.497)</td>
</tr>
<tr>
<td>F</td>
<td>15.370***</td>
<td>18.312***</td>
<td>5.348***</td>
<td>2.587*</td>
<td>20.755***</td>
<td>20.426***</td>
<td>5.348***</td>
<td>15.784***</td>
</tr>
<tr>
<td>(ANOVA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheffe test</td>
<td>(4)&gt;(1)</td>
<td>(5)&gt;(2)</td>
<td>(4)&gt;(1)</td>
<td>(5)&gt;(1)</td>
<td>(4)&gt;(2)</td>
<td>(5)&gt;(2)</td>
<td>(5)&gt;(1)</td>
<td>(4)&gt;(1)</td>
</tr>
<tr>
<td></td>
<td>(4)&gt;(2)</td>
<td>(5)&gt;(3)&gt;(1)</td>
<td>(5)&gt;(1)</td>
<td>(5)&gt;(2)</td>
<td>(5)&gt;(2)&gt;(1)</td>
<td>(5)&gt;(1)</td>
<td>(5)&gt;(1)</td>
<td>(4)&gt;(2)</td>
</tr>
<tr>
<td></td>
<td>(5)&gt;(2)</td>
<td>(5)&gt;(4)&gt;(1)</td>
<td>(5)&gt;(3)&gt;(1)</td>
<td>(5)&gt;(3)&gt;(1)</td>
<td>(5)&gt;(3)&gt;(1)</td>
<td>(5)&gt;(3)&gt;(1)</td>
<td>(5)&gt;(3)&gt;(1)</td>
<td>(5)&gt;(3)&gt;(1)</td>
</tr>
</tbody>
</table>

*p < 0.05 ***p < 0.001

### Discussion

NCGAS was developed that reliably measures three subscales (cognition, affection, and behavior) and five factors (learning, confidence, liking, participation and leisure) of students’ computer game attitudes. Scales formed from the items chosen to represent these constructs have good validity and internal consistency reliability. The instrument consists of an economic 22 items. It can be used as a core for larger questionnaires designed to address the role of other variables that impact on game-based learning, or it can be used in a stand-alone fashion to examine the impact of interventions intended to change these five factors and three subscales.

This study also revealed that male students have a more positive attitude toward using computer games for learning, demonstrate greater confidence in using computer games, enjoy these kinds of activities, and freely participate in these activities both in a classroom setting as well as in their leisure time at home. Male students earned higher scores in the Cognition, Affection and Behavior subscales than their female counterparts. These results complement those of previous studies indicating that males demonstrate very positive attitudes toward computer games while females show a less positive or neutral attitude on all the four subscales: affective components, perceived control, perceived usefulness, and behavioral components (Bonanno & Kommers, 2008; Chou & Tsai, 2007). Male students feel more strongly than females that computer games are useful learning tools. One possible explanation could be that females are always more skeptical than males about the instructional potential of computer games, considering the availability of other resources to provide what can be learned from computer games. In contrast, boys perceive computer games as a unique engaging experience, lacking entirely in other computer applications (Bonanno & Kommers, 2008). Another explanation could be that male students are more likely to embrace a manipulation-oriented information processing approach (Casey, 1996), a command strategy for executing tasks, and a competitive social comportment when using computer games (Singh, 2001; Rommes, 2002). With this in mind, computer games should be designed to satisfy these learning needs.
Second, the male students demonstrated greater confidence in controlling computer games by themselves. This is consistent with the findings of Bonanno and Kommers’ (2008) indicating that males feel much more in control of computer games, making them better able to perform demanding actions. The reason for this may be that boys tend to have more experience participating in computer games than girls, and such extensive exposure to computer games develops a higher level of confidence in using them (Brosnan, 1998). Success builds confidence; therefore, when playing a new game or difficult game level, females require on-going reinforcement and more guidance than males. In other words, female students may require more task-related support to build competence and confidence.

Third, male students reported deriving more enjoyment from the computer games than their female counterparts. This is in line with the study of Chou and Tsai (2007), and might be due to the presentation of game content. It appears that boys enjoy competition and gaining high scores; while girls prefer cooperation and completion. Kafai (1996) proved that most commercially-available computer games do not reflect the interests or tastes of girls. Therefore, a biography of gaming should be compiled and analyzed to identify the genres of computer games that make girls feel uncomfortable. Further studies could explore how girls perceive incompetence in computer gaming and self-exposure in collaborative computer game playing. Instructional computer game designs that are perceived as less threatening should be identified and introduced during the initial phase to support game based learning.

Finally, male students were more likely to participate in computer game related activities and play the computer game in their leisure time. This is consistent with the study of Chou and Tsai (2007) in which male students are motivated by the aspect of entertainment in playing computer games to a higher degree than female students are. One possible reason could be that females enjoy a greater variety of leisure activities, preferring those that are less intensely focused. This could be the reason that females feel less motivated to participate in computer game activities or play computer games in their leisure time. In contrast, playing computer games constitutes the main leisure activity for many male adolescents, contributing to the quality of their friendships (Griffiths & Hunt, 1995). For this reason, the findings of previous studies and the on-going habits of students should be considered when developing effective game based learning environments.

Gender differences were observed for all five factors, and male students showed stronger positive attitudes with regard to the Cognition, Affective and Behavior subscales. The most important contributing factor may be the stronger motivation of males to play computer games (Chou & Tsai, 2007). It is necessary therefore, to consider gender differences when dealing with the cognitive, affective, and behavioral aspects of incorporating computer games into school instruction. Female students should be provided continual reinforcement and support when participating in a computer game based learning environment.

The experience students have using the Internet was another variable to which researchers have paid considerable attention. This study revealed that the influence of Internet experience on the attitudes of students toward computer games parallels the findings in previous studies (Smith, Caputi, & Rawstorne, 2000; Tsai, Lin, & Tsai, 2001). Students with more experience using the Internet generally have a more positive attitude toward computer games. Our results indicate that prior experience with the Internet plays an important role in shaping attitudes toward computer games. Familiarizing students with the Internet may provide a practical means to shift the attitudes of students in a positive direction. Based on our empirical observations, many students use the Internet to play on-line games in their leisure time; therefore, the negative impact of computer games, such as Internet addiction disorder (Lin & Tsai, 2002; Tsai & Lin, 2003) or online aggression (Liu, Ho, & Song, 2011), should be noted when designing game-based learning environments.

Conclusion

The CGAS presented in a previous study (Chappell & Taylor, 1997) explains only 44% of all variables; however, our New CGAS has demonstrated higher validity, while covering a broader range of factors, most of which were developed in this study. Previous researchers addressing these issues tended to emphasize the affective component; however in this study, the investigation has been expanded to include components of cognition and behavior. In the past, gender differences in learning, confidence, participation, and leisure sub-scales were scarcely mentioned, while this study has provides a number of unique insights into this aspect of computer gaming. We have also initiated an investigation into the influence of Internet experience on NCGAS.
The results of this study largely parallel those of previous research addressing attitudes related to technology. Male students demonstrated greater enthusiasm for computer games than their female counterparts, as did all students with more experience using the Internet. Clearly, software developers should consider gender preference and Internet experience when developing educational computer games. The inclusion of female protagonists and suitable learning support for learners with various degrees of experience using the Internet would be good options. Computer games represent a fundamental element in the lifestyles of young adolescents, while learning environments and attitudes strongly influence the utilization of technology (Lee & Chen, 2008; Liu, 2010) Gauging the attitudes of students toward computer games is an important topic of research into game-based learning, and this study represents an important first step. The NCGAS is a valid and reliable tool, providing a comprehensive framework with which to gauge the attitudes of students toward computer games. Researchers involved in game-based learning could use this tool in collaborative learning environments to organize groups according to attitudinal preferences, particularly those found on the cognitive subscale: learning and confidence.

The next logical step in this investigation should focus on continued modification. Future researchers could develop additional attitude factors to address issues unique to computer games, such as addiction and aggression. This pilot study applied exploratory factor analysis to verify the structure of NCGAS, and we would encourage the use of confirmatory factor analysis (Chang et al., 2011; Tsai, Lin, & Tsai, 2001) to further examine the validity and reliability of the newly developed scale. Researchers are also encouraged to apply this scale in different countries and cultural backgrounds. In the future, NCGAS could be used to identify the attitudes of individuals of various ages and educational levels toward learning with computer games. A more extensive study should be conducted to reveal the most advantageous approach to using these games within existing educational systems. Subsequent studies could investigate the relationships among gaming attitudes, the structure of games, the setting of playing games, game types and educational objectives. A careful exploration of the attitudes of students regarding computer games in these areas could provide additional insight into the design of serious instruction based games.

Acknowledgements

This study was supported by the National Science Council, Taiwan, under contract numbers NSC 100-2631-S-008-001, NSC 100-2511-S-008-017-MY2 and NSC 100-2511-S-008-006-MY2.

References


Seamless Connection between Learning and Assessment- Applying Progressive Learning Tasks in Mobile Ecology Inquiry

Pi-Hsia Hung1, Gwo-Jen Hwang2*, Yu-Fen Lin1, Tsung-Hsun Wu1 and I-Hsiang Su1

1Graduate Institute of Measurement and Statistics, National University of Tainan, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // hungps@mail.nutn.edu.tw // gjhwang.academic@gmail.com // jkk1002@hotmail.com // tnwuts@mail.tn.edu.tw // d09611004@stumail.nutn.edu.tw

*Corresponding author

(Submitted April 11, 2011; Revised January 01, 2012; Accepted March 08, 2011)

ABSTRACT
Mobile learning has been recommended for motivating students on field trips; nevertheless, owing to the complexity and the richness of the learning resources from both the real-world and the digital-world environments, information overload remains one of the major concerns. Most mobile learning designs provide feedback only for multiple choice items to guide students' learning. The aim of this study was to develop a series of worksheets as scaffolding to support inquiry-based ecology observations in a mobile learning environment. Based on a three-layer worksheet framework, a mobile learning system with in-field scoring rubrics was developed to guide the students to sequentially focus on guided observations, independent observations, and extended inquiries. Well-designed worksheets, instant feedback and supplementary materials were provided to balance the students’ learning pressure and task challenge in the species-rich field. The automated scoring data for constructed response tasks were analyzed to investigate the learning growth and students’ learning characteristics. The results showed that the proposed approach was effective in improving the field observation performance of the students. Furthermore, nearly 70% of the students who learned with the proposed approach could pursue their own inquiries in the mobile learning environment, showing that the approach is promising.

Keywords
Mobile learning, Ubiquitous learning, Science inquiry, Inquiry-based learning, Growth assessment

Introduction
In the past decades, the constructivist paradigm has played an important role in educational research. From a constructivist perspective, it is crucial that learners are engaged in constructing their own knowledge in one form or another (Hmelo-Silver, Duncan, & Chinn, 2007; Mayer, 2004; Palincsar, 1998). The practices of scientific inquiry including posing questions, observing in the field, gathering, analyzing and interpreting data, and constructing evidence-based arguments aim to fulfill the objectives of the constructivist process (Kuhn et al., 2000; Krajcik & Blumenfeld, 2006). The purpose of education is to cultivate independent and self-directed learners (Bruner, 1985). However, the free exploration of a highly complex environment may generate a heavy working memory load that is detrimental for novice learners.

Cognitive and psychometric studies on human learning have revealed the potential to greatly influence instructional design principles (Gorin, 2006; Kalyuga, 2009). However, it is conceptually difficult for most teachers to apply effective guided instruction to cultivate self-directed learners by themselves (Cazden, 1988). The notion of scaffolding proposed by Wood, Bruner and Ross (1976) provides a good reference for coping with this problem. Many researchers have suggested that scaffolding can enhance comprehension, improve independent learning and application, and promote knowledge transfer (Greenfield, 1984; Bruner, 1985; Pea, 2004). In other words, by adaptively fading out the scaffolding, novice learners can gradually become independent learners (Palincsar, 1998; Mayer, 2004; Hmelo-Silver et al., 2007).

With the aid of advances in wireless communication and mobile technologies, mobile-learning (m-learning) has become a new trend in education, especially for those learning activities conducted in the field (Hwang, Kuo, Yin, & Chuang, 2010; Hwang, Shi, & Chu, 2011). Nonetheless, even though the educational application of wireless communication technology seems to be innovative and interesting, several issues have been revealed when devising instruction to assist students to learn in such a complex learning scenario. One major problem is how to take the
students’ knowledge levels into account in a novel and information-enriched situation which may impose additional cognitive load on learners’ working memory (Sweller, 1999; Sweller, 2005). Scaffolding helps to make the learning more tractable for learners by arranging different target tasks which are achievable, manageable, and within the student’s zone of proximal development (Quintana et al., 2004). Scaffolding can also decrease the cognitive load by guiding the novice learner to focus on aspects of the tasks that are directly relevant to the learning goals (Salomon et al., 1991; Hmelo-Silver et al., 2007).

Most of the current mobile learning designs provide online feedback only to multiple-choice items (e.g., Chen, Kao, & Sheu, 2003; Chu et al., 2010; Hwang, Wu, & Ke, 2011). To foster the in-field observation and inquiry competences of students, this study proposes an authentic mobile learning system with in-field scoring rubrics based on a three-layer worksheet design. The working memory load in this learning system was carefully adapted to match most students’ challenge levels. Accordingly, online scoring feedback for each task was developed to encourage students to deeply engage in learning. In other words, the automated scoring feedback was designed to help students effectively clarify the content knowledge, focus on the learning targets, and facilitate their inquiry progress. To evaluate the effectiveness of the proposed approach, an experiment was conducted to compare the learning performances of the students who learned with the mobile learning system and those who took the traditional in-field learning approach.

Literature review

Using mobile technology to facilitate indoor and outdoor learning has become a growing trend in education. Various studies have examined the use of such new technologies in school settings in an attempt to provide learners and educators with support that is more active and more adequate (Chen et al., 2003; Cavus & Ibrahim, 2009; Chen et al., 2009; Hwang & Chang, 2011). Previous studies have shown that mobile devices, such as PDAs (Personal Digital Assistants) or smartphones, can be used as a cognitive tool to efficiently provide information and feedback relevant to the current learning situation (Hwang et al., 2009; Hwang, Chu, Lin, & Tsai, 2011; Vogel et al., 2010). Consequently, learners will have more adequate mental capacity available to focus on specific learning tasks, which significantly improves student motivation and learning outcomes (Lai et al., 2007; Tan et al., 2007; Chu, Hwang, & Tsai, 2010). Furthermore, the portability feature that extends the learning environment from traditional classrooms to authentic contexts and appropriate cultures has been considered as being an important benefit of using mobile devices to learn (Juniu, 2002; Hwang, Tsai, & Yang, 2008; Hwang, Chu, Shih, Huang, & Tsai, 2010).

Even though mobile learning has been recognized as being an effective learning approach (Rogers & Price, 2009; Chu, Hwang, Tsai, & Tseng, 2010), there are few sound instructional design models in this area (Chiou, Tseng, Hwang, & Heller, 2010), especially with respect to the model of cognitive learning theory. Through synthesizing several mobile learning projects in the context of fieldtrips and outdoor experiences, e.g., Ambient Wood (Rogers et al., 2005), Environmental Detectives (Klopfé et al., 2002) and Savannah (Facet et al., 2004), Rogers and Price (2009) have suggested that one of the main challenges in mobile learning is to avoid information overload so that students are not too bewildered with an overly complex situation. Therefore, constant guidance provided by teachers, facilitators or learning systems, and explicit structure in the activities are key components for successful mobile learning (Scarlatos, 2006). Researchers have indicated that unguided or minimally guided instructional approaches could lead to less effective and efficient learning than guided instructional approaches owing to the ignorance of the structures that constitute human cognitive architecture (Sweller, van Merriënboer, & Paas, 1998; Kirschner, Sweller, & Clark, 2006). Furthermore, instructional design without solid rationales or theories may not take good advantage of new technologies.

Cognitive Load Theory (CLT) plays a major role in the theory of cognition and instructional design, and could provide an informative reference for developing such a sound learning environment (Sweller, 1999; Sweller, 2005). A key assumption of CLT is that human working memory capacity is limited, and that overloading working memory hinders learning (Sweller et al., 1998; Paas, Renkl, & Sweller, 2003). Accordingly, to facilitate learning, the working memory load needs to be carefully adapted to match the students’ challenge level. In terms of CLT, the intrinsic cognitive load is essential for achieving specific learning goals. The more elements and inter-relationships among elements that the to-be-learned material consists of, the higher the intrinsic cognitive load will be. In the outdoor ecology inquiry learning of this study, the students needed to use learning resources from both the real world and the digital world; moreover, there was a need to complete their observations during several field trips. Without a careful
learning design, the intrinsic cognitive load of the participants could be very high. In this study, two strategies were considered to help the students manage the cognitive load without exceeding the limits of their working memory capacity. First, the provision of various process worksheets (or worked examples) rather than a full problem to be solved helped the students break the learning activities into smaller tasks. Secondly, the separation of a three-layer worksheet consisting of different format and cognitive category tasks stimulated them to further foster in-field observation and inquiry competences.

Additionally, situated in an unfamiliar and information-enriched space, the students needed more direct information and feedback on a just-in-time basis to fill the gap between the complexity of the abundant to-be-learned material and their lack of basic knowledge about the ecosystem. To take the extraneous cognitive load into account, the PDAs used in this study played the role of a mobile e-library that provided corresponding supplementary materials during the field trips. This facility prevented the participants from spending too much time searching for information from the paper-based materials during or after the learning activities. As such, unnecessary information searching and processing was reduced, resulting in a lower extraneous cognitive load. In addition, the hints in the form of multiple-choice and short-answer questions provided step-by-step guidance to the students, so that they could focus on the information concerning the learning task they needed to work on at the time.

**Mobile learning environment with progressive learning tasks**

**Three-layer observation learning framework for mobile ecology inquiry**

This study proposes a three-layer framework for mobile learning (as shown in Figure 1) to balance students’ learning pressure and the challenges they face. The first layer uses multiple choice items and online feedback to clarify the students’ basic knowledge. The second layer provides short response items to help students focus on crucial observations. The students are encouraged to take notes and pictures for their inquiries. The third layer suggests that students answer their own questions and provide relational descriptions of the ecosystem to integrate and extend their conceptual knowledge.

![Figure 1. The three-layer observation learning scaffold design](image)

In addition, designed process worksheets provide sections for students to take notes and do diary writing and to raise further questions autonomously for the extended inquiry after the field trip. The scoring feedback is devised to motivate the students to further elaborate their records, reflections and inquiries. Students depend on their learning status to adaptively decide whether to immediately finish their work in the field or to catch up on all their missing data later on the web. Once the students’ unfamiliarity with the mobile devices and the outdoor learning activity is actually eliminated, they are able to access more cognitive resources to help them link the incoming information with their prior knowledge base, and to reflect on or monitor their performance. Ultimately, the design of the three-layer scaffolding, from guided observation, independent observation, to personal reflection, sequentially, stresses the need to gradually improve students’ competence in terms of contextualization, internalization of ecological knowledge, and reflective thinking.
The numbers of the different format and cognitive category tasks embedded in the observation activities are provided in Table 1. There are three different types of tasks (multiple choice, short response, and extended response) and two different cognitive levels (guided and independent observations). The multiple choice items are used to clarify the students’ basic knowledge. The adaptive feedback is provided online. Only short responses are required for the field observation tasks. The time-consuming extended response tasks are completed after the students get back to school.

Table 1. The task distribution of the mangrove observation worksheets

<table>
<thead>
<tr>
<th>Interface Item type</th>
<th>Level</th>
<th>Guided observation</th>
<th>Independent observation</th>
<th>Contrast and integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDA</td>
<td>multiple-choice</td>
<td>25</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>multiple true-false</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>short response</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>PC</td>
<td>extended response</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

The students can search for specific information from an e-library while they are working on the worksheets via the PDA, as shown in Figure 2. In the first-layer learning tasks, the learning system guides the students to observe the learning targets by asking a series of questions to clarify their factual knowledge. Based on their responses, the system then determines the next learning step, such as suggesting that they “observe the learning target again,” “search for more information from the e-library” or “go to the next task.”

Figure 2. The e-library interface

For the second-layer learning tasks, the students are encouraged to take notes on the note section of the PDAs about what they have observed autonomously in the field. In addition to the PDAs, each student is equipped with a telescope and a digital camera, with which they can observe the distant targets and take photos of what they have found to be interesting, as shown in Figure 3. After completing the collection of field observation data, the students then upload those records to the learning system through the wireless network.

Figure 3. Photos of learning activities in the field

After the observation activity, the students return to the computer classroom for the third-layer learning tasks. They can browse the data they have uploaded to edit their learning diaries with a personal computer; moreover, they are asked to extend the descriptions they have made in the field based on the notes and photos taken. They can also view
the teacher’s ratings and feedback on their worksheets while editing their learning diary. The purpose of editing the learning diaries is to provide opportunities for them to summarize and reflect on the field observations for further inquiries.

**In-field scoring rubrics**

There were two parts to the automated scoring mechanism: short responses on the PDA, and extended responses on the computer (as shown in Table 2). These scoring rules are developed based upon human rater scoring rubrics (Table 3). Human rater scoring rubrics were available online to communicate the major learning objectives. The first part of the automated scoring rules evaluates the students’ observation skills in terms of the quality and quantity of the observation responses on the PDA. For example, if a student provides 10 records on the PDA, he or she gets 4 points for the PDA section worksheet. The second part evaluates the student’s extended observation responses on the computer after the field observation activities. If a student provides 2 appropriate relationship descriptions in the diary, he or she gets 3 points. In this study, the students were all familiar with the scoring rubrics, because the instructors explained the scoring rules clearly at the very beginning of the field trips.

<table>
<thead>
<tr>
<th>rating elements</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responses in PDA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantity</td>
<td>10</td>
<td>6-9</td>
<td>3-5</td>
<td>1-2</td>
<td>no R.</td>
</tr>
<tr>
<td>accuracy</td>
<td>7</td>
<td>5-6</td>
<td>3-4</td>
<td>1-2</td>
<td>no R.</td>
</tr>
<tr>
<td>question</td>
<td>7</td>
<td>5-6</td>
<td>3-4</td>
<td>1-2</td>
<td>no R.</td>
</tr>
<tr>
<td><strong>Responses in diary</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>features</td>
<td>5 well described</td>
<td>5 relevant</td>
<td>3-4</td>
<td>1-2</td>
<td>no R.</td>
</tr>
<tr>
<td>relationship</td>
<td>3</td>
<td>1-2</td>
<td>1-2</td>
<td>incorrect</td>
<td>no R.</td>
</tr>
<tr>
<td>Inquiry</td>
<td>3 success</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>no R.</td>
</tr>
</tbody>
</table>

**Table 2. The scoring rubrics for the observation skills**

Figure 4 shows the architecture of the automated scoring system for the PDA-integrated learning environment. The feedback provided by the system includes scores, correct answers, explanations and references. The purpose of developing the system is to continuously sustain students’ efforts towards achieving the learning goals. A pattern-matching algorithm is employed to compare the similarity between the students’ descriptions with the sample description provided by the content expert. In each observation worksheet, three short response tasks are presented. For each task, the content expert was asked to provide some sample responses. For each response, there are three to eight keywords related to the learning task. The keywords and their synonyms for each sample response were then extracted and added to the keyword set of that task by the research team. During the learning activity, each student is
allowed to answer a question twice. The keywords in the student’s answers are compared with those in the keyword set of the task. A score of zero to four is assigned based on the number of matching keywords. The system also provides three different kinds of qualitative feedback, namely, “Good job! Your descriptions are correct and detailed.” for 3-4 points, “Correct descriptions, but more details are needed.” for 1-2 points and “Please search for supplementary materials in the e-library and modify your description.” for 0 points.

The accuracy of the automated scoring mechanism was evaluated by comparing its scores with those given by two human raters. Table 3 shows an example of contrasting the human scoring rubrics and automated scoring rules for a constructed response task. Raters apply scoring rubrics to rate students’ responses, while the scoring system employs the pattern-matching method based on a pre-defined keyword set. In this example, the learning task is “Describe the physical features and eating behaviors of male fiddler crabs” and the keyword set is “white body, a large front claw, a small claw, mud, and organic.” In this study, the inter-rater reliability between the two raters for the three tasks and the total scores are .94, .91, .90, and .93, respectively. For the scoring consistency between the human raters and the computer system, the correlation coefficients for the tasks range from .82 to .93, revealing that the automated scoring has reached an acceptable consistency with the human rating.

It should be noted that the objective of providing feedback in such in-field inquiry-based activities is quite different from that in web-based learning activities. In a mobile or ubiquitous learning activity, the main learning content consists of those real-world learning targets in the field. It is improper to provide complex feedback or learning content with rich information to interfere with students' in-field observations, in particular, on the mobile devices with small screen size. Therefore, in this study, only necessary feedback, such as the scores and short responses that reflect individuals' present performance, is provided.

| Table 3. A contrasting example of human scoring rubrics and automated scoring rules |
|----------------------------------------|-----------------|----------------|----------------|
| Human rating                          | Sufficient descriptions of physical features and eating behaviors. | Descriptions of two required features but only one of them is sufficient. | Description of one feature only or descriptions of two features partially. | Descriptions are irrelevant to the task. |
| Automated scoring                     | 4-5 keywords    | 2-3 keywords   | 1 keyword      | 0 keywords     |

**Experiment design**

To evaluate the applicability of our approach, an experiment was conducted in the Chiku mangrove wetland located in the southwest coastal region of Taiwan for a science curriculum activity. The wetland is the largest and the most intact lagoon in Taiwan. The main water source of the wetland is Zengwun River which brings large varieties of biological, nutritive products, thus forming an excellent habitat for wildlife. Many shorebirds, wildfowl and egrets gather at sandbars in the river, including the famous endangered “Black-faced Spoonbill.” The aim of this learning activity was to foster the students’ observation, description, and inquiry competences, which are considered as important educational objectives of natural science courses in Taiwan; therefore, the learning activity was designed as part of the distinguishing curriculum of selected elementary schools.

**Participants and learning procedure**

The subjects were forty-nine 5th and 6th graders aged from 11-12 years old. Twenty-four students were assigned to the control group and twenty-five students were the experimental group. The experimental group experienced the field trips within four months in the Chiku mangrove wetland and learned with the mobile learning system proposed in this study. On the other hand, the control group learned with traditional inquiry-based approach; that is, they were guided by paper worksheets to obtain information about the ecological system from the teacher's presentations, the web and the in-field observations.

To ensure that the students were able to focus on the major learning objectives while making explorations in the field, an orientation was given to introduce the learning tasks. The orientation included a whole picture of the
learning environment and the learning tasks. The students in the experimental group further received a 40-minute lesson on some basic practices regarding the use of the PDAs.

For the first trip, the students in the experimental group spent about three hours completing the observation tasks with the worksheet. For the second trip, the students spent two hours visiting an ecology pond near Chiku. They were encouraged to observe and raise questions for further studies. The third trip was a particularly interesting outing for the students. They took a boat to observe the ecology system of a green tunnel for 30 to 40 minutes. They then spent another 2 hours clarifying their own puzzles arising from their previous observations or extending their own inquiries in this trip. They also used digital cameras to collect data, and telescopes for long-distance observations during the learning activities. After each field trip, the students were asked to go back to the computer classroom to complete a learning diary by searching for relevant information from the Internet, making reflections upon their experiences in the field and reorganizing what they had recorded in the field.

Two experienced teachers were then asked to rate the worksheets and learning diaries submitted by the students. Finally, the students took a questionnaire concerning their attitudes toward the mobile learning activity for field observations. In addition, a reward system was used to balance the students’ learning motivation and fun experiences during the fieldtrip. The students were told that they would receive a certificate if they completed the learning tasks; moreover, the five students with the highest scores would receive a gift.

**Instrument**

The science inquiry ability assessment developed by Lin (2009) was applied to assess students’ inquiry abilities. The assessment consisted of three facets, i.e., retrieving information, interpreting information, and reflecting on and evaluating content of on-line texts. There were a total of 12 items in the assessment. Its alpha coefficient was .90. The correlation between the assessment and the school science grade was around .40, showing acceptable reliability of the assessment.

**Data collection design**

To evaluate the improvements in the students’ learning performance, the automated scoring results of the students in the first field trip and the third field trip are compared. The scoring aspects included were the students’ observation records, questions raised, descriptions of the learning targets, and the surrounding environment. Those records were also rated by two teachers to check the validity of the automated scoring results. The consistency coefficients of the human and machine ratings range from 0.82 to 0.93. In addition, a questionnaire was used to collect the experimental group students’ feedback on using the PDAs while participating in the field trips, including their attitude toward using the PDAs to learn in the field, their feelings about the pressure of using the PDAs to learn in the field, the usefulness of the e-library, and the helpfulness of the PDA learning system. The questionnaire consists of four items with a four-point Likert scale and one open-ended question.

**Results and findings**

**Science inquiry ability assessment of the two groups**

The science inquiry ability assessment was used as the criterion variable. One-way analysis of covariance was applied to examine the effects of different learning conditions. The results indicated that the experimental group's post-test score of the science inquiry ability assessment was significantly better than that of the control group (8% variances accounted by group variable, as shown in Table 4). That is, the proposed mobile learning system could substantially promote the students’ learning performance in in-field inquiry-based learning activities in comparisons with the traditional approach.

The lower growth of the control group on the inquiry abilities (from 0.49 to 0.58 in mean) could be attributed to a lack of scaffolding for reducing their extraneous cognitive load and increasing their germane cognitive load that can stimulate learners to engage in more meaningful inquiry learning (Sweller, 1999; Sweller, van Merriënboer, & Paas,
The finding revealed several possible problems in traditional inquiry-based approach, including the insufficiency of in-field experiences during the learning process and the lack of instant guidance and feedback for cultivating students' higher order thinking abilities in fields (Hwang, Wu, Tseng, & Huang, 2011).

Table 4. Analysis of ANCOVA of the on-line science inquiry abilities for two groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Mean (Adjusted)</th>
<th>SD</th>
<th>F</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>25</td>
<td>0.35</td>
<td>0.69</td>
<td>0.94 (0.97*)</td>
<td>0.81</td>
<td>4.72*</td>
<td>0.08</td>
</tr>
<tr>
<td>Control</td>
<td>24</td>
<td>0.49</td>
<td>0.68</td>
<td>0.58 (0.55#)</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. # adjusted mean   * p < .05

Table 5 further shows the t-test results of the in-field performance of the experimental group students on the first and the third trips. The results suggest that they made substantial progress in every aspect, including their observation records, the questions raised, and their descriptions of the learning targets and the ecology environment. By analyzing the embedded worksheet data, we can conclude that the scaffold design successfully helped them to enhance their ecology inquiry skills. This internal checking design is direct, valid, efficient, and economical in terms of achieving the learning objectives. Whenever students submit their responses, the scoring data will be ready for analysis. Using automated scoring data, the students’ learning progress can be evaluated by objective statistics without any extra costs.

Table 5. The t-test results of the extended responses between trips 1 and 3

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation records made</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on the trip</td>
<td>Trip 1</td>
<td>2.72</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>Trip 3</td>
<td>4.08</td>
<td>1.26</td>
</tr>
<tr>
<td>Questions raised on the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trip</td>
<td>Trip 1</td>
<td>0.76</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Trip 3</td>
<td>2.28</td>
<td>0.84</td>
</tr>
<tr>
<td>Descriptions of learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>targets in the learning</td>
<td>Trip 1</td>
<td>4.48</td>
<td>1.09</td>
</tr>
<tr>
<td>diary</td>
<td>Trip 3</td>
<td>5.32</td>
<td>1.44</td>
</tr>
<tr>
<td>Descriptions of ecology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relationships in the</td>
<td>Trip 1</td>
<td>2.36</td>
<td>0.70</td>
</tr>
<tr>
<td>learning diary</td>
<td>Trip 3</td>
<td>2.72</td>
<td>0.46</td>
</tr>
<tr>
<td>Questions raised in the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>learning diary</td>
<td>Trip 1</td>
<td>1.04</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Trip 3</td>
<td>3.20</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01, N = 25

Outcomes of the students with different performance levels

According to the learning portfolios of the experimental group students, the teachers categorized them into three performance levels after a 30 minute discussion. The performance levels were mainly determined by the accuracy of the students’ descriptions and the number of questions they raised. The advanced students used more scientific vocabulary to describe what they had observed, and provided more accurate and detailed descriptions; in addition, most of the questions they raised were scientifically oriented. On the contrary, the descriptions submitted by the basic level students contained more inaccurate content with less scientific vocabulary; moreover, the questions raised were shallow. As shown in Table 6, about half of the students were classified as being at the proficient level, implying that those students were good at describing the main features of their observations, although some of their descriptions could be insufficient or inaccurate. Moreover, nearly 28% of the students reached the advanced level; that is, they were able to use scientific vocabulary to accurately describe what they had observed in detail.

Table 6. The means of the short responses for different level students

<table>
<thead>
<tr>
<th>Level</th>
<th>Quantity</th>
<th>Accuracy</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced (28%)</td>
<td>3.14</td>
<td>2.86</td>
<td>3.00</td>
</tr>
<tr>
<td>Proficient (44%)</td>
<td>2.55</td>
<td>1.36</td>
<td>2.00</td>
</tr>
<tr>
<td>Basic (28%)</td>
<td>1.43</td>
<td>0.57</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Note. N = 25
Table 7 shows the extended responses of the different performance level students in their learning diaries. It was found that the pattern of their performance was similar to that shown in Table 6. However, for those basic level students, the scores in Table 6 are noticeably lower than those in Table 7; that is, providing responses immediately in the field is a challenging task for those basic level students since they need to face a complex learning scenario that consists of real-world learning targets as well as the learning task instructions or supplementary materials from the learning system at the same time. For the learning reflection part, the differences between levels are also smaller, except in the extended inquiry learning. The basic level students made few autonomous inquiries (0.86) since they raised few questions. More stimulating examples may be needed to help them to raise some scientific-oriented questions. On the other hand, the advanced students outperformed the others in every aspect; that is, they were quick to capture the critical features of the observed targets, and their descriptions were detailed, accurate, and related to their own experiences.

There were significant differences among the different level students. The characteristics of the different levels in terms of their observation, description and reflection skills are summarized as follows. For the basic level students, the number of valid short or extended responses was limited (around 2); the observation records were few and incomplete; their observation descriptions did not focus on the learning targets; and the extended responses in their learning diaries were shallow. For the proficient level students, the number of valid short responses was around 5, while the number of valid extended responses was around 3; most of their descriptions were relevant and scientific oriented; and about 1 autonomous question was successfully pursued. For the advanced level students, the quality and quantity of the observation descriptions were both good; the number of short responses and extended responses were around 7 and 4 respectively; most of their descriptions were relevant and accurate; and about 2 autonomous questions were pursued successfully.

The basic level students might have put too much effort into observing species or using the technology devices to search for information such that they could not cope with the mobile ecology inquiry context. Therefore, the number of questions they raised was significantly lower than the number raised by the advanced level students. In other words, the students’ cognitive load played a key role in their learning progress. Even though the online worksheet and scoring feedback could partially satisfy the basic level students’ need for support for knowledge and concept learning, their questioning competence was still too weak to allow them to pursue their own inquiries.

<table>
<thead>
<tr>
<th>Level</th>
<th>Observation</th>
<th>Reflection</th>
<th>Inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced (28%)</td>
<td>2.71</td>
<td>2.14</td>
<td>2.00</td>
</tr>
<tr>
<td>Proficient (44%)</td>
<td>1.82</td>
<td>2.09</td>
<td>1.36</td>
</tr>
<tr>
<td>Basic (28%)</td>
<td>1.71</td>
<td>1.71</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Note. N = 25

Question 1: Describe the physical features, living environments and special records of “Kandelia obovata” in the field. Your descriptions will be scored based on the aspects of quantity, quality, accuracy and relevance.

<table>
<thead>
<tr>
<th>Category</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical features</td>
<td><em>Rhizophoraceae</em> family, a small evergreen tree, five meters in height. It has grayish-brown bark. Its leaf has opposite leaf arrangement and is long oval. The leaf has thick skin to prevent loss of water. <em>It’s a viviparous plant. Its pen-like viviparous seedling stretches out from circular cone-shaped mature fruit. The green viviparous seedling turns reddish-brown and falls into the muddy ground when mature.</em></td>
</tr>
<tr>
<td>Living environments</td>
<td>The living area is between river and sea in tropics or subtropics.</td>
</tr>
<tr>
<td>Special records</td>
<td><em>The color of bark is grayish-brown. The base of the tree significantly thickens into a plank shape. The dense prop roots branch out into soft mud from the tabular root.</em></td>
</tr>
</tbody>
</table>

Figure 5. An example of an advanced level student’s performance
A sample of an extended response of an advanced level student is given in Figure 5. This record includes 3 features pertaining to the leaf, 2 features of the fruit, 3 features of the body or branch of the tree, and one feature related to the growing environment. The automated scoring result can easily assess that the student performed well. He or she was able to describe the features of the observed plants as well as their surrounding environment. With appropriate guidance in the field, nearly a quarter of the participants were able to provide such detailed descriptions of their field observations in this study. Most of this level’s students are ready to pursue their own inquiries autonomously.

**Students’ perceptions of the mobile ecology observation learning system**

At the end of the third trip, a questionnaire was used to collect feedback from the students in the experimental group about their attitudes toward and perceptions of the mobile observation activity. Twenty-two students completed the questionnaire. It was found that 82% of the participants showed positive attitudes toward using the PDAs in the observation activities; for example, several students expressed that “I am more willing to participate in outdoor learning activities than before.”

Moreover, 91% of the students felt that using mobile devices to learn in the field could reduce their learning pressure in comparison with the traditional approach. For example, one student indicated that “I feel that it is easier and less pressured when using the PDA system to learn in the field in comparison with the traditional approach.” Other students also expressed similar feelings. Around 95% of the students indicated that using the PDA system to learn had made them concentrate more on the learning targets than they would in the traditional approach.

On the other hand, only 50% of the participants agreed that the e-library was helpful to them during the learning activity. This information suggests that more efforts need to be made in training the students to search for and use relevant information in the e-library. For example, it could be helpful to extend the 30-minute e-library introduction and searching practice in future applications.

**Conclusions**

In this paper, a three-layer mobile learning worksheet design with in-field scoring rubrics was developed to assist students in identifying their learning targets, focusing on critical details of observation, and extending their own inquiry learning. The results suggest that proposed mobile learning approach could substantially promote the students’ in-field inquiry performance in comparisons with the traditional approach. The students’ perceptions of the learning system further suggested that the learning and assessment design proposed in this study was workable and enjoyable for most of the participants. Moreover, the collaborating teachers thought that the proposed mobile learning approach successfully engaged the students in the outdoor ecology inquiry activities in a spiral manner; that is, the three-layer worksheets and the corresponding scoring feedback helped the participants at different performance levels focus on their observation tasks and raise their own questions for ecosystem knowledge construction. Such a learning design provides an opportunity for elementary school students to conduct scientific observations in a well-supported mobile learning environment.

Although the experiment results and feedback from the students have demonstrated that our approach is applicable and effective for integrating mobile learning systems into outdoor learning designs, it should be noted, however, that about 28% of the students did not exhibit good performance. Their observation descriptions were rather shallow and even irrelevant to the learning tasks, which could be due to their lack of prior knowledge, barriers in understanding the objective of the learning tasks, or their lack of observation or communication ability. To provide more effective assistance to such students, it is important and challenging to investigate in depth the factors that might affect the learning performance of such students, and develop adaptive mobile learning mechanisms by taking those factors into consideration. If sufficient time is allowed, most students can make reasonable progress in worksheet tasks. In other words, the time constraint in the field may negatively affect basic level students’ performance. Thus the time distribution issue should be considered in future practice (Hwang, Kuo, Yin, & Chuang, 2010). Furthermore, it would be of great value to integrate the mobile learning system into the regular science curriculum (Hwang & Tsai, 2011; So, Kim, & Looi, 2008; Wong, Chin, Tan, & Liu, 2010). In addition, the continuous assessment data from the automated scoring system could also be valuable for describing the change in the students’ attitudes and conceptual structure. An online scoring mechanism for constructed response tasks is a promising direction for future mobile
inquiry research; in the meantime, more rigorous experimental data are needed in the future for clarifying the feedback effect in this mobile learning and assessment design model.

Acknowledgements
This study is supported by the National Science Council of Taiwan under contract numbers NSC 99-2511-S-011-011-MY3 and NSC 100-2631-S-011-003.

References


A Maturity Model for Assessing the Use of ICT in School Education

Mauricio Solar1, Jorge Sabattin2 and Victor Parada2

1Universidad Técnica Federico Santa María, Chile // 2University of Santiago de Chile, Chile // mauricio.solar@usm.cl // jorge.sabattin@usach.cl // victor.parada@usach.cl

(Submitted April 11, 2011; Revised September 26, 2011; Accepted November 25, 2011)

ABSTRACT

This article describes an ICT-based and capability-driven model for assessing ICT in education capabilities and maturity of schools. The proposed model, called ICTE-MM (ICT in School Education Maturity Model), has three elements supporting educational processes: information criteria, ICT resources, and leverage domains. Changing the traditional and exclusive focus on ICT, five Leverage Domains are defined: Educational Management, Infrastructure, Administrators, Teachers and Students. The Leverage Domains generate a hierarchical structure with a second level named Key Domain Areas. These areas should be measurable and controllable, so they are related to a third hierarchical level, called Critical Variables, allowing the model’s elements to be assessed qualitatively and quantitatively. The capability and maturity of these variables associated with the intersection with the other two elements establish five levels of capability. The proposed model is strongly supported by the international standards and best practices for ICT management. It has been validated through data collection instruments and its associated web–support tool was also refined with a small pilot study. In summary, the proposed ICTE-MM model provides a basis for self-assessment and improvement planning. It is not just a diagnostic tool but has also been found to be useful for guiding the principals in ICT investment.

Keywords

Evaluation methodologies, Secondary education, ICT in education, Maturity model

Introduction

Different studies indicate that several parameters considerably affect the implementation of new technology in the school environment (Condie & Munro, 2007; Kozma, 2008; Zamani, 2010; Underwood et al, 2010). For example, the opinions and attitudes of principals with regard to Information and Communication Technology (ICT) adoption influence the extent and quality of technological applications’ usage (Pelgrum, 2001). In fact, two schools with the same infrastructure, same human resources and same students can have very different results. There are concrete cases in which a mere change of the school’s principals has generated, with the same teachers, striking results in a short time (Waissbluth, 2010). Given this, principals and administrators need a vision to help them towards their goals. It is especially important at the school level for the principal to have a vision of what is possible through the use of ICT, and to be able to work with others to achieve that vision (Sun, Heath, Byrom, Phlegar, & Dimock, 2010).

As an example, consider the following very well-established technology vision of a school (Nido, 2011):

“The purpose of technology use at school is to support teaching and learning. Technology is a powerful learning tool that, when properly integrated into a challenging curriculum, improves learning and helps us achieve our educational goals.

Students are active learners who use technology responsibly to solve problems, develop critical thinking skills and communicate ideas. Teachers integrate technology across the curriculum where appropriate to enhance instruction and assessment of student learning. They take advantage of staff development designed to improve teaching and learning with technology. Technology provides parents with user-friendly access to up-to-date and meaningful information about the school and their children’s academic status. The school administration is committed to improving the efficiency of school technology by providing current and effective resources, appropriate training and qualified personnel.”

Every school needs an ICT plan to achieve its vision, but according to Zhihua and Zhaojun (2009), evaluation is usually the weakest component of ICT plans. There are several factors involved, but an important one is that policy makers often have unrealistic expectations about the learning improvements that will result from ICT initiatives. For example, when you consider that it takes an average of four to five years for most teachers to reach a level of
technological proficiency at which they can use computers fluidly and effectively, then an impact on student learning will not occur any sooner.

On the other hand, the degree of success that a school has in implementing ICT will depend, in part, on the quality and maturity of its ICT plan. Therefore, the school needs to monitor its progress towards ICT integration and use the evaluation results to plan its ICT program, e.g., identifying needs, problems, and opportunities as well as specific aspects of the program such as professional development, technical assistance, and resources. Evaluation is neither easy nor inexpensive, but when it is an integral part of the ongoing ICT planning and infusion process, it is well worth the effort (Spector, Merrill, van Merrienboer, & Driscoll, 2007).

It is clear that schools need a periodical evaluation mechanism to know where they are. This mechanism also indicates how to advance (more rapidly) towards the correct direction (roadmap) without losing time, efforts or resources. Evaluation is necessary, but unfortunately there are no standardised approaches. The choice of an evaluation method would depend on what aspect of ICT and education we want to evaluate; consequently, there is a wide range of assessment tools in ICT and education. An example of these proposals is that from Rodriguez, Nussbaum, López and Sepúlveda (2010). These authors present a monitoring and evaluation scheme for a specific ICT for an education program that supports teaching and learning using mobile, computer-supported collaborative learning. Using the information provided by the program, the authors analyse the impact of the innovations adopted by teachers on student achievement.

The development of capability maturity models has been a strong trend in various technological and organisational areas. These models are proving to be useful because they allow individuals and organisations to self-assess the maturity of various aspects of their processes against benchmarks. The best-known models are those belonging to the Capability Maturity Model Integration family developed by Humphrey at Carnegie Mellon University (CMMI, 2006). These models are typically constructed with five levels, where each maturity level provides a new foundation of practices on which subsequent levels are built. Although they were developed for the software products and services, their capability maturity level structure and the mechanisms for determining those levels have been replicated by many other models in other areas (Valdes, Solar, Astudillo, Iribarren, Concha, & Visconti, 2011). Some applications of the maturity model concept in e-learning are detailed in Harris (2004), Neuhauser (2004), and Marshall and Mitchell (2006). All these proposals were developed as an E-learning Maturity Model to provide a means by which tertiary institutions (universities and polytechnics) can assess and compare their capability to develop, deploy and support e-learning (focused on learners rather than teachers and institutions). An example of these applications in school education is shown in Zhihua and Zhaojun (2009). They present a maturity model framework applied during the ICT Test Projects in Shanghai rural schools, providing appropriate structures to deliver effective educational experiences through ICT use. The framework includes six models: “Technology,” “Curriculum,” “Leadership/Management,” “Workforce,” “Inter/intra-institutional linkage” and “External linkage.” Each model consists of a number of dimensions, 63 dimensions in total for the six models. Each of the dimensions within these models may be treated on a six-point Likert scale with positive scoring (Likert, 1932); a school that satisfies all the attributes within a given level of a dimension is assigned an appropriate score. The authors suggest that the models could be used as an action guideline for school managers, one that would prevent deviation from the scheduled targets. The maturity model of Zhihua and Zhaojun (2009) is not based on international standards.

A maturity model should be developed based on the internationally recognised standards of ICT use in education (Bonina & Frick, 2007). It should be adjusted to the given school’s local conditions and measure the degree of compliance with the internationally recognised standards (Olsson, 2006). It allows one to diagnose the situation, and from there generates a roadmap that guarantees a virtuous cycle of continuous improvement. The roadmap helps to optimise the ICT investments and allows the school to reach higher levels of maturity.

The proposed maturity model suggests a way to measure the use of ICT standards in the school. This model, based in CMMI (2006), is a process improvement approach that provides organisations with the essential elements of effective processes. The model will be called "Maturity Model for ICT in School Education" (ICTE-MM). The following sources have been considered as inputs in designing the proposed model:

- Models that provide the standard structure of a CMMI (2006).
- ISTE's National Educational Technology Standards (NETS), the NETS for Teachers (NETS-T, 2008), the NETS for students (NETS-S, 2007), and the NETS for Administrators (NETS-A, 2002).
Technology Standards for School Administrators (TSSA, 2001).

The ICTE-MM model was validated with 19 stakeholders through different data collection instruments, a complementary self-assessment web tool was pilot-tested with six schools, and a fine-tuned version was generated that incorporates the participants’ feedback and an ICT, implementation roadmap for each evaluated school.

The proposed ICTE-MM model allows one to evaluate a school against the best international practices of educational ICT use, including formulation of organisational strategies, management of ICT, operative management, and capabilities of the school and its human resources (teachers, students and administrators). The next section establishes the problem statement and the general structure of the model. The section “Logic of Proposed Model: Leverage Domains and Key Domain Areas” presents the details of the model’s components as well as describes a rationale for the model. We then explain the validation process of the model and a pilot study of the web tool. In the last section, we conclude by discussing the main advantages of the model.

**Problem statement**

The research done by Zhihua and Zhaojun (2009) suggests that schools that are more e-mature improved their performance levels significantly and more quickly than those that are not. Saiti and Prokopiadou (2009) observe that the usage of new technologies in school administration is vital to upgrade administrative processes. Similarly, Marshall and Mitchell (2006) conclude that the results of the application of the e-learning maturity model have been found valuable for strategic planning activities in universities. In general, the e-maturity model of Zhihua and Zhaojun (2009), as well as Underwood et al. (2010) proposal, tries to answer the question: “Does maturity model have a positive impact on learner outcomes?”. The purpose of our proposal is specifically to answer the following questions: “Does ICTE-MM support principals to make appropriate decisions on ICT investment? Does ICTE-MM support principals to identify how to develop future actions?

According to Waissbluth (2010), Chilean principals have non-professional faculties to lead, and a consensus has not developed around basic aspects, as much administrative as pedagogical, of the management and leadership of organisations. In the Chilean context, Hepp, Hinostroza, Laval and Rehbein (2004) note that “introducing ICT into the schools, without a proper staff development plan and without a pedagogical perspective, is a low-return investment.” Where resources are limited, this lesson really must be learnt quickly (Wagner, Day, James, Kozma, Miller, & Unwin, 2005).

Public schools with limited resources have no budget to hire an in-house expert exclusively for technical assistance. Several studies (Eteokleous, 2008; Saiti & Prokopiadou, 2009) indicate that ensuring the quality and permanency of technical assistance constitutes one of the important factors for the efficient introduction and exploitation of new technological capabilities.

All these evidences suggest that principals need support to make decisions about ICT investments. In this paper we propose a maturity model in order to provide principals with virtual technical assistance (roadmaps) to make appropriate decisions on whether to expand or modify a particular policy or ICT investment and to identify how to develop future actions.

**Research context**

The study investigated computer use in the schools of Santiago, Chile. The population of Santiago is about 6 million. Nationally, there are 10,605 schools distributed among four categories according to their financial sources: 6,250 public schools (59%), 3,217 subsidised schools (30%), 1,068 private schools (10%) and 70 schools of entrepreneurial corporations (0.1%). There are 2,576 schools, equivalent to 24.29% of the national total, of which 611 are public, 1,615 are subsidised, 317 are private and 33 are entrepreneurial corporations. As for the number of students, the total enrolled at the regional level in 2007 was 1,405,200 (Mineduc, 2009).
Figure 1. Hierarchical structure of (5) domains, 25 KDA and CV
Proposed Maturity Model for ICT in School Education (ICTE-MM)

Following the ISTE's National Educational Technology Standards (NETS-A, 2002; NETS-S, 2007; NETS-T, 2008) and TSSA (2001), we defined five Leverage Domains: Educational Management, Infrastructure, Administrators, Teachers and Students (Figure 1). The Leverage Domains generate a hierarchical structure of levels. The second level is named Key Domain Areas (KDA). These areas should be measurable and controllable and are related to a third hierarchical level called “Critical Variables” (CV). Each KDA can be measured by whether it meets its goals, which are determined by the CV that determine the given KDA’s capacity.

The “Information Criteria.” To satisfy the school requirements, the information must satisfy certain criteria which constitute the ICTE requirements for this information. These criteria for the information provided by a given KDA are:

- **Effectiveness**: the information must be relevant and pertinent as well as being delivered timely, correctly, and consistently.
- **Efficiency**: the information must be generated by the most productive and economical use of resources.
- **Confidentiality**: the information must be protected from unauthorized disclosure.
- **Integrity**: the information must be accurate and complete.
- **Availability**: the information must be available when required by the stakeholders, and its associated resources and capabilities must be safeguarded.
- **Compliance**: the information must comply with those laws, regulations and contractual arrangements to which the school is subjected, i.e., externally imposed educational criteria as well as internal policies.
- **Manageability**: information must be easy to deal with and usable by management to operate the school.

The “IT Resources.” The Leverage Domains require “IT Resources” to generate, store and deliver the information required to reach the school objectives. IT resources are:

- **Applications**: information systems and manual procedures used to process data and generate information.
- **Data**: on every format required by the school and processed by the information systems.
- **Infrastructure**: technology (e.g., hardware, operating systems, database management systems, networking, multimedia, etc.) that enables the processing of the applications.
- **Facilities**: the environment that houses and supports the IT infrastructure.
Logic of proposed model: Leverage domains and key domain areas

Leverage Domains are the model’s core elements because they are used to establish the capability levels, which are in turn compared with the current status of a given school. Five leverage domains and 25 KDA (Figure 1) were defined by studying international standards along with input and feedback provided by instructional technology experts and educators, including classroom teachers, administrators, teacher educators, and curriculum specialists (Tomei, 2005). We detail two domains, Management and Infrastructure, including the description of their critical variables. Due to limited space, the other three domains, Administrators, Teachers, and Students, are described in a general context (see Teachers and Students’ variables in Figure 1).

Management

This domain is based on Technology Standards for School Administrators (TSSA, 2001). It provides criteria to evaluate the ability of the school to articulate an ICT consistent vision. In other words, the school should have an ICT strategy aligned with the school’s overall strategy, and there should be an explicit commitment to the ICT. The management domain’s KDAs are as follows:

**(MAN-1) School Management.** This domain allows for the verification and qualification of plans in accordance with school administration, specifically for the measurement and monitoring of the following six CVs: action plan for ICT use in administration; technology plan in administration; monitoring and Evaluation plan; verification of the personal use of ICT in administrative positions; generation of report cards, registration, qualifications (i.e., SIMCE and PSU, nation-wide school evaluations in Chile); and follow-ups of student/teacher cases (record of educational cases for students and teachers).

**(MAN-2) Vision, Strategies and Policies.** This domain allows for the management of all ICT resources according to the school’s vision, strategy and priorities. At the same time, it aligns the plan with national educational policies. It specifically allows for the measurement and monitoring of the following CVs: alignment strategies; commitment of senior management; communication with the school community; resource allocation (policy for the use of internet resources; and policy for the acquisition of inputs and technological resources meeting school needs).

**(MAN-3) Organisation and ICT Management.** This domain outlines activities helping the school to organise and manage ICT. It is evaluated through the following CVs: planning guidance IT infrastructure; IT infrastructure planning; defining the organisational structure; and IT process roadmap.

Infrastructure

This domain is based on the maturity model described in Valdés et al. (2011). It provides guidance on how the school can develop its multimedia resources to provide the foundation for ICT implementation. This domain contains the following KDAs:

**(INF-1) Software.** This measure allows for the development of infrastructure-oriented application; it is defined within the context of the school’s educational objectives and its integration with other internal and external systems. It is evaluated through the following CVs: operating system; educational software; and Administrative Software.

**(INF-2) Networks.** This domain measures the network architecture that defines the communications infrastructure for the transmission of information. Its CVs are the following: internet; wi-fi; and intranet.

**(INF-3) Hardware.** This measure allows for the development of an ICT infrastructure that defines the technologies and standards for the technical components that enable the ICT initiatives. The variables are as follows: access to the computer room; quality of technological equipment for educational use; access to equipment deployment information multimedia; computers available for education; and access equipment information capture.

**(INF-4) Maintenance Plan.** This domain allows for measurement planning that maintains the operability of ICT in the school. Its CVs are as follows: maintenance of equipment; operational maintenance supplies; and presence of a maintenance plan.
(INF-5) Security. This measure allows for the development of a security infrastructure that defines the technologies and safety standards to ensure that internal and external transactions are secure. It is measured with the following CVs: condition safety; insurance contracts; health conditions; personnel for security work; and backup information.

Administrators

These standards are indicators of effective leadership for ICT in schools based on NETS-A (2002):

(ADM-1) Leadership and Vision. Educational leaders inspire a shared vision for comprehensive integration of ICT and foster an environment and culture conducive to the realisation of that vision.

(ADM-2) Learning and Teaching. Educational leaders ensure that curricular design, instructional strategies, and learning environments integrate appropriate ICT to maximise learning and teaching.

(ADM-3) Productivity and Professional Practice. Educational leaders apply ICT to enhance their professional practice and to increase their own productivity and that of others.

(ADM-4) Support, Management, and Operations. Educational leaders ensure the integration of ICT to support productive systems for learning and administration.

(ADM-5) Assessment and Evaluation. Educational leaders use ICT to plan and implement comprehensive systems of effective assessment and evaluation.

(ADM-6) Social, Legal, and Ethical Issues. Educational leaders understand the social, legal, and ethical issues related to ICT and model responsible decision-making related to these issues.

Teachers

This domain is based on the international standard NETS-T (2008). Teachers design, implement, and assess learning experiences to engage students, improve learning, enrich professional practice and provide positive models for students, colleagues, and the community. All teachers should meet the following standards and performance indicators. This domain’s KDAs are as follows:

(TEA-1) Student Learning and Creativity. Teachers use their knowledge of the subject matter, teaching and learning practices, and ICT to facilitate experiences that advance student learning, creativity, and innovation in both face-to-face and virtual environments.

(TEA-2) Digital-Age Learning Experiences and Assessments. Teachers design, develop, and evaluate authentic learning experiences and assessments, incorporating contemporary tools and resources to maximise content learning in context and to develop knowledge, skills, and personal attitudes.

(TEA-3) Digital-Age Work and Learning. Teachers exhibit knowledge, skills, and work processes representative of an innovative professional in a global and digital society.

(TEA-4) Digital Citizenship and Responsibility. Teachers understand local and global societal issues and responsibilities in an evolving digital culture and exhibit legal and ethical behaviour in their professional practices.

(TEA-5) Professional Growth and Leadership. Teachers continuously improve their professional practice, model lifelong learning, and exhibit leadership in their school and professional community by promoting and demonstrating the effective use of digital tools and resources.

Students

Within a solid educational framework, ICT can help students to live, learn and work successfully in an increasingly complex society that is rich in information and knowledge. Students and teachers should use digital technology effectively. This domain’s KDAs are based on the following international standard NETS-S (2007):
(STU-1) Creativity and Innovation. Students demonstrate creative thinking, acquire knowledge, and develop innovative products and processes using ICT.

(STU-2) Communication and Collaboration. Students use digital media to communicate and work collaboratively, at times long-distance, and to support individual learning and contribute to the learning of others.

(STU-3) Research and Information Fluency. Students apply digital tools to gather, evaluate, and use information.

(STU-4) Critical Thinking, Problem Solving, and Decision Making. Students use critical thinking skills to plan and conduct research, manage projects, solve problems, and make informed decisions using appropriate digital tools and resources.

(STU-5) Digital Citizenship. Students understand human, cultural, and societal issues related to ICT and practice legal and ethical behaviour.

(STU-6) Technology Operations and Concepts. Students demonstrate a sound understanding of technology concepts, systems, and operations.

Capability model and maturity determination

The capability is a measurement of the state of each KDA that contributes to support the school’s development. The capability of a KDA is determined by using the Capability Level (CL) of each of its Critical Variables (CV); in other words, the capability of these variables to satisfy certain requirements is evaluated. The capabilities of the critical variables are averaged to give a final KDA CL.

To accommodate ICTE strategies with different variables relevance, weights (W) are used for each variable group. Thus, the CL of a KDA is the weighted average of the CLs of its variables CV (Eq. 1).

\[ CL_{KDA} = \text{Average}[\text{CAP}(CV_1) \times W_1, \text{CAP}(CV_2) \times W_2, \ldots, \text{CAP}(CV_n) \times W_n] \]

(Eq. 1)

The weights \( W_i \) used in this first model application are all equal, but the model allows to adjust their values according to the needs.

Maturity is a property of the school as a whole, and the maturity level (ML) is obtained from the KDA capacity levels that the school has (\( CL_{KDA} \)). There are several options to determine an organization maturity, namely:

1. Minimum CL among all KDAs
2. Average CL of all KDAs
3. Predetermined KDA configuration, using a set of values for all KDAs in model.
4. Configuration of high-priority KDAs, using a set of minimum values for all KDAs in the model.

The last criterion (Configuration of high-priority KDAs) was adopted in ICTE-MM. The school ML corresponds to a predetermined configuration of KDAs in predefined CLs (Eq. 2). See Table 1.

\[ ML = \text{Conf}1(CL_{KDA1}, \ldots, CL_{KDAk}) \]

\[ ML = \text{Conf}5(CL_{KDA1}, \ldots, CL_{KDA6}) \]

(Eq. 2)

Table 1. Example of organizational ML with a set of high-priority KDAs

<table>
<thead>
<tr>
<th>Domain</th>
<th>KDA</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management</td>
<td>School Management</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Vision, Strategies &amp; Policies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization &amp; ICT Management</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Software</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Network</td>
<td></td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hardware</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maintenance Plan</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Security</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Administrators</td>
<td>Leadership &amp; Vision</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This mechanism was selected for ICTE-MM for its flexibility to allow graduating progress according to specific ICTE strategies, since it only requires to fix a minimum set of KDAs that are important for a given ML; development criteria and rates for other KDAs are left to the school. The actual criteria to use can be extracted from domain specialists or school leaderships. Once the current ML is assessed, the model allows one to identify the states required to advance to a higher level and to propose a “roadmap” to improve the school.

The model considers five levels for a staged evolvement of KDA capability. Each KDA includes variables with capability levels of their own; a weighted average of the variables’ capability levels determines the KDA capability level. A set of common patterns was identified to determine the discrete capability levels. For instance, level 1 means that capability does not exist, although the school may have recognised the importance for such capability; in level 2, the capability exists, but it is neither structured nor formalised; in level 3, the capability exists and it is well documented and structured; in level 4, the capability is structured, and its metrics and automatic tools have been defined and standardised to improve its effectiveness and efficiency; finally, level 5 implies all of the capabilities above plus the use of best practices and international standards in its achievement.

**Description of KDA capability levels**

As a sample of how KDA capability levels are defined, the capability model of one KDA is described in Table 2 (School Management). As defined in the “Management” section, “School Management” allows for the verification and qualification of plans in accordance with school administration.

<table>
<thead>
<tr>
<th>Critical Variable</th>
<th>Level 1 “Initial”</th>
<th>Level 2 “Developing”</th>
<th>Level 3 “Defined”</th>
<th>Level 4 “Managed”</th>
<th>Level 5 “Optimised”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action plan of ICT use in administration</td>
<td>There is no plan</td>
<td>There is an informal plan</td>
<td>There is a formal plan</td>
<td>There is a formal and standardised plan</td>
<td>There is a plan, and it is annually reviewed</td>
</tr>
<tr>
<td>Technology plan in administration</td>
<td>There is no plan</td>
<td>There is an informal plan</td>
<td>There is a formal plan</td>
<td>There is a standardised plan</td>
<td>There is a plan based on the use of best practices</td>
</tr>
<tr>
<td>Monitoring and evaluation plan</td>
<td>There is no plan</td>
<td>The plan is informal</td>
<td>There is a formal plan</td>
<td>There are some automatic tools</td>
<td>There is a plan based on the use of best practices</td>
</tr>
<tr>
<td>Verification of the personal use of ICT in administrative positions</td>
<td>There is no verification</td>
<td>There is an informal verification</td>
<td>There is a formal procedure</td>
<td>There is a standardised verification process</td>
<td>There is a procedure based on the best practices</td>
</tr>
</tbody>
</table>
The KDA “Vision, Strategies and Policies” must satisfy the ICTE objective of managing and conducting all “IT resources” according to the school strategy and its priorities. Its relevant “Information Criteria” are Effectiveness and Availability, and the main “IT Resources” required are Applications and Data.

Its level of capability is determined by six critical variables, but we show details of the following four variables: (1) Strategy alignment with the national educational directions. (2) Principal commitment with the implementation of school initiatives. (3) Periodic communication to all involved people within the school. (4) Resource assignment commitment with the implementation of the organizational school strategy.

The CLs are defined below. Within each level four assertions are presented, one for each variable related to the KDA.

- **Level 1 “Initial”**: (1) There is evidence that the school has recognized that the strategy alignment is important and needs to be addressed; however, there are neither actions nor approaches that tend to be applied. (2) There is no awareness and need for the principal to get involved early with the school initiatives. (3) There are no formal actions to communicate the school initiatives to the people in the school. (4) There is no evidence of resources specifically allocated for the ICT and education implementation.

- **Level 3 “Defined”**: (1) The ICTE Vision is well defined and it is integrated to the school strategy. There is a policy about ICT and educational strategy planning and it is well documented. (2) Principal and teachers are committed to and get involved early in the ICTE initiatives. (3) The ICTE vision, policies and strategies have been communicated to and are well understood by all personnel in the school. (4) Enough monetary resources to support ICTE initiatives have been assigned. Its allocation is included in the school’s annual budget.

- **Level 5 “Optimised”**: (1) The vision is periodically reviewed according to stakeholders’ needs and new technologies. The strategy and policies are periodically updated according to feedback from stakeholders, teachers, and educational policies. The strategy planning process is continuously compared with the international standards. (2) Principals have an explicit role assigned in the IT strategy planning process. (3) Personnel, stakeholders, and teachers are considered when the ICTE vision is developed. (4) Resources assigned to the ICTE initiatives are periodically adjusted according to a cost/benefit analysis and to stakeholders satisfaction.

**Validation of the ICTE—MM model**

The evaluation methodology associated to the ICTE-MM model establishes activities, schedules, institutions, workflows, work products, roles and responsibilities, to provide effective assessments to schools using the ICTE-MM model (Sabattin, 2009). The evaluation methodology is supported by a web tool, which supports the gathering and processing of information generated in each evaluation based on the ICTE-MM model.

Table 3 shows the participants and data collection instruments used during the validation process. In order to check validity we use more than one data source to obtain information about key aspects of the model: The proposed model has been presented to instructional technology experts (2), teachers (6), administrators (5), and principals (6). These recipients represented schools of different maturity levels (according to annual ICT investment), sizes (number of students) and types (Pub: public; Sub: subsidised; Pri: private). A survey was conducted to determine the relevance of the CV included in the ICTE-MM model, and a questionnaire was administered to determine whether the model was a valid and useful tool for guiding principles in moving towards best management practices. The questionnaire
was sent to 19 stakeholders as shown in Table 3, and all of them (100%) answer it in less than 5 days. So, according to Margaret Sanger Center International (2009), the results are representative of the group of stakeholders to whom the questionnaire was mailed. The results show that 80% of the CVs were considered of high or very high relevance. Most remarkable is the case of the “Teachers” domain, which yielded a relevance perception of 92%. Due the current model weights of critical variables are equal, all participants (100%) agreed that future versions of the model should refine these weights.

Table 3. Participants in validation process

<table>
<thead>
<tr>
<th>Technology experts</th>
<th>Teachers</th>
<th>Administrators</th>
<th>Principals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pub</td>
<td>Sub</td>
<td>Pri</td>
</tr>
<tr>
<td>Number of participants</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Data collection tools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaires</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Interviews</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

During interviews, we assured the principals that what they were going to say was in confidence and that the identity of speakers would not be revealed. As participants were guaranteed anonymity, we can ensure they were honest in their responses (research validity). Since there are only three interviews, all three were conducted by one experienced and trained interviewer, avoiding invalidity in comparisons of data across different interviewees. The interviewer received training about the project and his role, and he received an interview protocol to conduct the interview. The training includes evaluation objectives, review of data collection techniques, and the importance of following instructions and the application protocol.

The principals (88%) stated that they agreed with the criteria for moving from one maturity level to the next. Five of the six principals found that a tool like this could be very helpful. The remaining one, a principal of a private school, did not need a tool like this because the school had hired an in-house expert for technical assistance.

Pilot study of the web-support tool

A pilot study conducted in a set of schools was performed during September 2009 with the following objectives: validate the model and its associated web-support tool; test the applicability of the model to schools with different characteristics; and obtain feedback from schools.

A sample of six heterogeneous schools was selected in terms of type, organisational size, and degree of technological advancement. All of them are based in the Metropolitan area of Santiago, Chile. The KDA set was divided up and assigned to three different management field roles: Administrator (director of the school, not always a teacher, but sometimes with a Ph.D. in education), IT Management (head of IT unit, normally a teacher), and Teacher. Therefore, at least three teachers were required to participate in each school.

A two-hour workshop was conducted to explain the web tool, a call center was available to answer questions during ten days and questions were answered by email as well.

The ICTE-MM model was validated and commented on by participants in the pilot study. This pilot study was supported by a web tool that supported the gathering and processing of information generated in each evaluation. The web tool was developed in PHP, using PhpMyAdmin and MySql to run in Joomla. We created six accounts to access the web tool, and we sent usernames and passwords to each participant. The web tool is available for any school and can be periodically used to monitor improvement of capabilities and maturity levels at the website http://tice-mm.inf.santiago.usm.cl (in Spanish).

An assessment is performed on a school by answering the web tool’s questions. The result is an evaluation (capabilities of each KDA and school maturity) plus a proposed capability maturity improvement roadmap. The pilot test was helpful to find out how long data collection takes: participants took ten days to answer all the questions of the web tool. They validated the model’s structure, applicability, and capability and maturity calculation scheme.
Furthermore, the feedback we received enabled us to make improvements to the structure of the model. As a result, we moved two CVs from one domain to another one.

Conclusions

The main benefits identified in the ICTE-MM model application are:

- It is a reference framework to identify the areas that support an ICTE strategy, based on internationally recognized standards for ICT management and education.
- The proposed ICTE-MM model provides a basis for self-assessment and improvement planning. In terms of theoretical implications, it is not just a diagnostic tool but has also been found to be useful for guiding the principals in moving towards best practices in management and ICT investment.
- The Leverage Domains, along with their KDA and variables that determine the capability maturity, are involved in a dynamic cycle, the school requirements acting as the main triggering input. This cycle allows for each school to achieve continuous improvement of ICT use in education, making it easier for them to transit towards higher maturity levels through planed roadmaps. One of the most important practical implications of the model is that the application of the ICTE-MM provides data to formulate technological projects and to base budgetary requests tailored to the individual school. This may help the principals to base their decision making on facts.
- It has a methodology to determine an improvement roadmap; besides the 5 leverage domains, 25 KDA and 103 CVs, the model incorporates organizational capabilities and maturity. Each KDA critical variable has minimum required values for higher MLs, marking improvement roadmaps for the school.
- The developed model was built and tailored to the characteristics of public schools in a developing country. This target school has a low budget and no in-house expert for technical assistance. Consequently, the ICTE-MM model is operationalised with a web tool for self-assessment of critical variables that acknowledge the capabilities of each school to carry out ICTE initiatives and guide principals in technical decisions.

Finally, this model distinguishes between capacity, a characteristic of KDA, and maturity, as a property of the school as a whole. The level of maturity allows one to define a "roadmap" to improve the school’s ability to address the challenges of education in the country.

Acknowledgments

This work was partially funded by DGIP-UTFSM, Project No. 241142 and the Electronic Government Center of the Informatic Department - UTFSM.

References


Student’s Uncertainty Modeling through a Multimodal Sensor-Based Approach

Imène Jraidi and Claude Frasson
Université de Montréal, Department of Computer Science and Operations Research, 2920 chemin de la tour, H3T-138 QC, Canada // jraidiim@iro.umontreal.ca // frasson@iro.umontreal.ca

(Submitted May 1, 2011; Revised September 25, 2011; Accepted December 1, 2011)

ABSTRACT
Detecting the student internal state during learning is a key construct in educational environment and particularly in Intelligent Tutoring Systems (ITS). Students’ uncertainty is of primary interest as it is deeply rooted in the process of knowledge construction. In this paper we propose a new sensor-based multimodal approach to model users’ uncertainty from their affective reactions and cognitive and personal characteristics. An experimental protocol was conducted to record participants’ brain activity and physiological signals while they interacted with a computer-based problem solving system and self-reported their perceived level of uncertainty during the tasks. We study key indicators from affective reactions, trait-questionnaire responses, and individual differences that are related to uncertainty states. Then we develop models to automatically predict levels of uncertainty using machine learning techniques. Evidence indicated that students’ uncertainty is associated to their mental and emotional reactions. Personal characteristics such as gender, skill level, and personality traits also showed a priori tendencies to be more or less in particular uncertainty states. The SVM algorithm demonstrated the best accuracy results for classifying students’ uncertainty levels. Our findings have implications for ITS seeking to continuously monitor users’ internal states so they can ultimately provide efficient interventions to enhance learning.

Keywords
Student uncertainty, EEG, Physiological sensors, Affect, Intelligent tutoring systems

Introduction
Research in distance education, and more precisely in Intelligent Tutoring Systems (ITS), tended to privilege cognitive aspects of teaching in which the learning process was considered as a set of information processing steps devoid of affective aspects, until studies in cognitive science, artificial intelligence, and neuroscience show that the brain mechanisms associated with emotions are not only related to cognitive mechanisms (Cytowic, 1989), but also solicited in perception, problem solving, decision making, and other cognitive processes (Cytowic, 1989; Damasio, 1994; Picard, 1997). Since then, various research areas including education, psychology, computational linguistics, and artificial intelligence devote a growing interest in the close links between affect and learning (Breazeal, 2003; Conati, 2002; Lester, Towns, & FitzGerald, 1999; Picard, 1997) as emotions have an impact on attention, motivation, memorization, and information processing (Goleman, 1996; Pekrun, 1992).

Affective user modeling has become a key construct in human-computer interaction and particularly in ITS (Conati & Maclaren, 2005; Picard, 1997). Even if there is no validated universal theory on emotions neither any consensus or agreement about which emotional states are related or pertinent to learning (Picard et al., 2004; Woolf et al., 2009), some studies rely on theoretical models explicitly linking emotions to learning (D’Mello et al., 2008; Kort, Reilly, & Picard, 2001) while other research focus on particular states such as frustration (Burleson, 2006; McQuiggan, Lee, & Lester, 2007), stress (Prendinger & Ishizuka, 2005) or attention (Rebolledo-Mendez et al., 2009).

Students’ uncertainty is of primary interest as it is considered among the most recurrently observed states during computer tutoring and due to its theorized relationship to learning (Craig et al., 2004; Graesser & Olde, 2003; Kort et al., 2001; Pon-Barry et al., 2006; VanLehn et al., 2003). Indeed, the state of uncertainty is deeply rooted in the process of knowledge construction; it is related to a state of confusion or hesitation that one inevitably goes through when a misconception or a lack of knowledge or understanding arises. It can also signal a lack of confidence with regards to ones self-efficacy in performing specific tasks. Depending on the case and the frequency of this state, a tutor can decide either to intervene by providing appropriate aid or hints to help the student to clear up his misconception, encouraging him to be more confident in the learned concepts, or to let him get over this state by himself.
Hence, a tutoring system has to efficiently identify the student’s state in order to formulate the appropriate response and adapt his pedagogical strategies accordingly. Nevertheless, actual ITS still cannot compete with human tutors who can readily detect from a glance that a student appears uncertain. In most work so far, uncertainty modeling relied on acoustic-prosodic, lexical, or discourse features extracted from utterance/dialogue-based system interactions (D’Mello et al., 2008; Liscombe, Hirschberg, & Venditti, 2005; Pon-Barry et al., 2006). However, uncertainty has barely been linked to learners’ emotional and mental manifestations. We consider that this state inevitably involves these dimensions, and that current discourse/utterance features can be insufficient or even imprecise, as they cannot always reflect users’ uncertainty. We also believe that uncertainty encompasses as well, cognitive factors and is specific to each individual and context.

In this paper we propose a new multimodal sensor-based approach to model learner uncertainty by integrating these factors. We use various data sources from learners’ electrophysiological activity assessing their mental and emotional reactions as well as cognitive and personal characteristics. The hypotheses we establish is that (1) these features are related to the state of uncertainty and (2) can effectively predict a learner’s uncertainty level. An experimental study was conducted to test these hypotheses. In this experiment, an acquisition protocol was established to record student affective reactions, trait-questionnaire responses and individual differences while they interacted with a logical problem solving system and expressed various levels of uncertainty.

Our approach is two-fold: First we analyze key trends that are associated to uncertainty, then we develop predictive models to automatically assess uncertainty states, which involves training machine learning algorithms using reported levels of uncertainty to supervise the classification process. Evidence indicated significant correlations between the electrophysiological sensor data and students’ reported uncertainty levels. Personal characteristics such as gender, skill level, and personality traits also showed a priori tendencies to be in particular uncertainty states. The SVM algorithm demonstrated the best accuracy results for classifying students’ uncertainty levels.

This paper is structured as follows. In the first section we present related work on automatic detection of students’ uncertainty and sensor-based affective modeling in the tutoring community. In the second section we describe our experimental setup and methodology. In the third section we present and discuss the obtained results. We conclude in the fourth section and present future work.

Related work

Promising results have been reported on correlating uncertainty to learning (Craig et al., 2004; Graesser & Olde, 2003; Kort et al., 2001; Pon-Barry et al., 2006; VanLehn et al., 2003). VanLehn et al. (2003) view uncertainty as a “learning impasse” that occurs when students realize that they lack knowledge, getting thus more involved to understand the material they are learning and about which they are uncertain. This creates an opportunity for the student to engage in constructive learning. Grasser and Olde (2003) describe uncertainty as a “cognitive disequilibrium” in which learners confront difficulties that fail to match their expectations, which causes deliberation and inquiry aimed at restoring cognitive equilibrium. Other studies show that adapting and responding to student uncertainty can greatly improve learning (Forbes-Riley & Litman, 2010; Pon-Barry et al., 2006).

Besides significant studies have been conducted on automatically recognizing uncertainty in tutoring systems (Carberry & Schroeder, 2002; D’Mello et al., 2008; Liscombe et al., 2005; Pon-Barry et al., 2006). Pon-Barry et al. (2006) for example, used linguistic cues (such as hedges, response latencies, or filled-pause signals) extracted from human tutoring corpus through a frequency analysis to detect users’ uncertainty in a tutoring system. Liscombe et al. (2005) used acoustic-prosodic features to classify student uncertainty in a corpus collected from a speech-enabled intelligent tutoring system. Carberry and Schroeder (2002) proposed an algorithm to recognize doubt by examining linguistic and contextual features of dialogue in conjunction with world knowledge including stereotypical beliefs ascribed to the dialogue. We believe, however that uncertainty is a rather complex state, which inevitably involves an affective dimension manifested by particular mental and emotional activations and is specific to each individual and context and that current discourse/utterance features can be insufficient or even imprecise, as they cannot always reflect users’ uncertainty. The main contribution of this study is to propose an alternative approach for uncertainty recognition based on new information sources such as users’ affective reactions and personal characteristics.
On the other side, the integration of physiological data combined with artificial intelligence techniques in ITS proved their effectiveness in assessing user state, trying to bridge the gap between actual tutoring systems and face to face education and improve technology’s adaptability by accurately detecting student’s affect, adapting tutorial interventions, and providing appropriate strategies to assist him to foster optimal conditions for learning (D'Mello et al., 2005; McDaniel, et al., 2007; Picard, 1997; Prendinger & Ishizuka, 2005; Woolf et al., 2009). Most of these studies use non-intrusive sensors to analyze a variety of physical cues including observable changes like face expressions, body postures, vocal tones, and physiological signal changes such as heart rate, skin conductivity, temperature, or respiration.

Moreover, with the advent of consumer oriented electroencephalograms (EEG), it is now possible to measure a learner’s mental state with a high time resolution and precision and develop systems that directly modulate their tasks to neural indexes of cognition. The growing progress in developing portable, convenient, and low cost EEG headsets and devices allows using EEG technology within operational educational environments (Chaouachi, Jraidi, & Frasson, 2011; Stevens, Galloway, & Berka, 2007). Neural research established various EEG-based mental gauges of alertness, engagement, or executive load using features extracted from power spectral density (PSD) bands or event related potential (ERP) components (Pope, Bogart, & Bartolome, 1995; Prinzel, Freeman, & Scerbo, 2000; Sterman et al., 1993). More precisely, EEG studies on mental concentration and attention defined an EEG indicator of attention to internal processing during performance of mental tasks (Harmony et al., 1996). They have found that an increase of the brain activity within the delta and low theta frequency band is related to an increase in subjects’ mental concentration.

In this paper, we propose a new multimodal sensor-based approach to model students’ uncertainty by integrating affective indicators using neurological and physiological sensors to track users’ emotional activations and mental concentration as well as cognitive and personal criteria within a problem solving context. We seek to identify key trends/indicators that are related to uncertainty states and develop a predictive model to assess students’ uncertainty levels.

**Methodology and experimental design**

The experimental setup consists of a problem solving system, a 6-channel EEG headset, physiological sensors, and two video feeds. Data were synchronized using necessary time markers in order to integrate the recorded signals with the rest of the instrumental setup under specific (un)certainty states. The problem solving system consists of a series of logical tasks that do not require particular perquisites but involve a high level of attention. These tasks imply inferential skills on information series and are typically used in brain training exercises or tests of reasoning. The system is composed of 3 modules. Each module is concerned with specific forms of data: the first module deals with geometrical shapes, the second module with numbers, and the third module with letters. Each module starts with a tutorial explaining the task and giving examples to get users accustomed with the types of problems. Then, 5 multiple-choice questions related to each tutorial are given. Learners were asked to respond as quickly and efficiently as possible. They were informed that a correct answer is rewarded 4 points, -1 point is given for a bad answer, whereas 0 point is given for a no-answer. A fixed time limit of 80 seconds for each question was imposed. Failing to give an answer within the allowed time was considered as a no-answer. We detail our methodology and protocol in the following subsections.

**Considerations for uncertainty elicitation**

One of the most important points in this study was to obtain accurate data related to specific uncertainty states. Thus problem tasks were selected in a way that potentially causes uncertainty. To choose the right answer, learners needed to deduce a logical rule. Without this rule, the learner was not able to be sure of his answer. Moreover, problems had different difficulty levels and some of them involved a second rule to decide between two answers that both match the first rule. For instance in the geometrical module, three shapes were successively presented in the interface. The first shape represented a black triangle, the second a white rectangle and the third a black pentagon. The learner was then asked to deduce the fourth element by choosing one answer among five possibilities. In this example, the rule that one should deduce is to add a side in each shape and the correct answer would be a hexagon. Two hexagons (black and white) were included among the propositions and only one matches to the second rule that one should also
deduce (i.e., alternating between the two colors) and the correct answer would be the white hexagon. Other questions were designed to systematically mislead the learners. For instance in the number-based module, two perpendicular data series were presented. In the vertical series all the numbers were multiples of seven and in the horizontal series all the numbers were multiples of five. In this task, one should deduce the missing intersection element, which should be a multiple of both five and seven. But no such element was given among the possible answers.

After each question, the system interacted with the learners and prompted them to report how they answered to the question by choosing between the following: “I was certain about my response” or “I was uncertain about my response.” Furthermore to assess uncertainty granularity levels, learners were prompted to choose between the following: “I was certain at 50% or more” or “I was certain at less than 50%”, whenever an uncertain response was reported. Hence three possibilities can be registered for each question: certain (Cert), uncertain (Uncert) and no-answer (No_Resp) with two possible granularity levels for Uncert, namely certain at 50% or more (Low_Uncert) or certain at less than 50% (High_Uncert).

Electrophysiological recordings

Three types of sensors were used during the experiment namely electroencephalogram (EEG), skin conductance (SC), and blood volume pulse (BVP) sensors. Data were digitized using the ProComp Infinity multi-channel data acquisition system (Thought Technology Ltd., 2007).

EEG is a measurement of the electrical brain activity produced by the synaptic excitations of neurons. During the session, learners wore a stretch electro-cap and EEG was recorded from sites P3, C3, Pz, and Fz as defined by the international 10-20 electrode placement system (Jasper, 1958). Each site was referenced to Cz and grounded at Fpz. Two more active sites were used namely A1 and A2 typically known respectively as the left and right earlobe. This setup is known as the “referential linked ear montage” and is depicted in figure 1. In this montage, roughly speaking, the EEG signal is equally amplified throughout both hemispheres. Moreover, the “linked-ear” setup yields a more precise and cleaner EEG signal by calibrating each scalp signal to the average of the left and right earlobe sites. For example, the calibrated C3 signal is given by \((C3 - (A1 + A2) / 2)\). Each scalp site was filled with a non-sticky proprietary gel from Electro-Cap and impedance was maintained below 5 Kilo Ohms. Any impedance problems were corrected by rotating a blunted needle gently inside the electrode until an adequate signal was obtained. The recorded sampling rate was at 256 Hz. Due to its weakness (on the order of micro volts), the EEG signal needs to be amplified and filtered. Besides, the electrical brain signal is usually contaminated by external noise such as environmental interferences caused by surrounding devices. Such artifacts alter clearly the quality of the signal. Thus a 60-Hz notch filter was applied during data acquisition to remove these artifacts. In addition, the acquired EEG signal easily suffers from noise caused by user body movements or frequent eye blinks. Thus a 48-Hz high pass and 1-Hz low pass de-noising filters were applied.

BVP and SC sensors were placed in the resting left hand fingers. Data were recorded at a sampling rate of 1024 Hz. SC measures changes in the resistance of the skin produced by the sweat gland activity. A tiny voltage is applied through two electrodes strapped to the first and middle fingers on the palm side. This establishes an electric circuit and allows us to quantify the skin's ability to conduct the electricity. BVP sensor was placed on the tip of the ring.
finger. It emits an infrared light and measures the amount of light reflected by the surface of the skin. This amount varies with the amount of blood present in the skin and thus with each heartbeat.

Affective data gathering

From the EEG raw signals, we computed mental concentration. As previously mentioned, this neural index is given by the brain activity within the delta and low theta (delta_low_theta) frequency band (Harmony et al., 1996). An EEG power spectrum was calculated for each electrode site using a Fast Fourier Transformation and the needed frequency band was extracted (1.56 - 5.46 Hz). We then computed a relative power value from the transformed signal by calculating the rate of the delta_low_theta sub-band range over the total EEG frequency band range (1.56 – 48 Hz). EEG relative power values were then summed from the electrode sites P3, C3, Pz, and Fz to compute the global ratio. A mean relative power band rate was measured for each task of the logical test.

SC signals were used to derive the galvanic skin response (GSR) widely known to linearly vary with the arousal ratings (Lang, 1995). It increases as a person becomes more stressed. From the BVP signal, the heart rate (HR) was calculated by measuring the inverse of the inter-beat intervals (distance between successive pulse peaks). The HR is extensively applied to understand the autonomic nervous system function and has shown a close correlation to valence (Lang, 1995). Both mean HR and mean GSR values were recorded for each entry. Normalization was done by subtracting current values from the baseline, and dividing the difference by the standard deviation.

HR and GSR were jointly used to measure specific emotional activations as emotions can be characterized in terms of judged valence (negative to positive) and arousal (low to high) (Lang, 1995). We used Russell’s Circumplex model of emotions (Russell, 1980) that classifies emotions within the two-dimensional arousal/valence emotional space. Two strategic emotional regions were defined during learning as depicted in figure 2. The first region involves negative emotions like frustration, boredom, or anger (negative region I and II) and should be avoided. The second region is the target emotional region specified by a slight positive valence and neutral arousal. This region is assumed to provide a maximum of efficiency and productivity in learning (Kaiser, 2006). In our study, we focused on the proportions of positive emotions within the target region for each question. We weighted then the number of HR and GSR recordings corresponding to this region by the total number of recordings.

Participants and protocol

Thirty-eight learners (14 women) with a mean age of 27.31 ± 6.87 years ranging from 19 to 47 years were recruited for the experiment. Participation was compensated with 10 dollars. Upon arrival at the laboratory, participants were briefed about the experimental objectives and procedure and asked to sign a consent form. Learners were then outfitted with the sensors and a 5-minute baseline was recorded to establish a neutral state for the electrophysiological parameters. Problem solving tasks were then completed and response time was recorded for
Learners were then asked to fill in information about their age, gender, skill level in logical based problem solving (low, or medium to high), and scales on a personality test, namely the Big Five Inventory (BFI). This test scales personality traits according to five dimensions, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN) (John, Naumann, & Soto, 2008).

**Results and discussion**

Our main hypotheses were that users’ brain activity, emotional reactions, and cognitive and personal characteristics (1) are related to their uncertainty level, and (2) can be effectively used to predict their actual uncertainty state. Figure 3 shows the general architecture of our approach. After completing the recording process (described in the previous section), we first identify through correlational analyses, key indicators from the recorded data that could be associated to uncertainty (I). Then we develop predictive models to detect levels of uncertainty using machine learning algorithms (II).

Figure 3 shows the general architecture of our approach. After completing the recording process (described in the previous section), we first identify through correlational analyses, key indicators from the recorded data that could be associated to uncertainty (I). Then we develop predictive models to detect levels of uncertainty using machine learning algorithms (II).

A total of 570 entries (15 questions x 38 participants) were gathered from this experiment: 323 for Cert responses (56.67%), 189 for Uncert (33.16%) with 103 entries for High_Uncert (54.50%) and 86 Low_Uncert (45.50%), and 58 No_Resp (10.17%). We detail in the following subsections the results obtained from both analyses.

**Key trends in learners’ uncertainty**

We started by investigating the relationships between learners’ uncertainty and their affective reactions according to the electrophysiological signals recorded across the four response types of the problem solving questionnaire. Statistical testing was performed using one-way analyses of variance (ANOVA). Figure 4 shows the results for the means of delta_low_theta relative power rates, HR, and proportions of emotions within the target region of Russel’s Circumplex model of emotions.

First, a main effect of response type was found for the delta_low_theta relative power values. An although small but significant difference was observed across the four conditions ($F(3, 566) = 3.559, p < 0.05$). This suggests that a statistically significant difference of mental concentration (signaled by the rates of the delta_low_theta power band) exists between the four reported levels of uncertainty. We observed the highest rates of delta_low_theta for the No_Resp and the High_Uncert groups of answers. This suggests that a state of certainty (i.e., the learner is sure of his reasoning and hence his response) does not necessarily imply a higher level of mental concentration and that being uncertain does not mean a lack of mental concentration but can be instead a sign of that if we also consider that in the No_Resp, learners were indeed very uncertain and did not take the risk to respond so they do not lose one point from the final score of the quiz, or that they did not find the answer within the allowed time. One can explain
this from another perspective, i.e., in case of uncertainty, the learner tends to be more focused and involved, trying harder to reach the solution of the problem and having difficulty in finding the logical rule between the data, which costs him a higher level of mental concentration as opposed to a state of certainty in which he is more at ease with the exercise. This confirms previous studies about the theorized relationship between learning and uncertainty (Craig et al., 2004; Graesser & Olde, 2003; Kort et al., 2001; Pon-Barry et al., 2006; VanLehn et al., 2003) where it is suggested that uncertainty can signal the advent of constructive learning, since that students tend to be more engaged to understand and clarify the fuzzy knowledge and concepts causing their uncertainty.

Second, a statistically significant difference between the types of responses was found for the HR signals ($F(3, 566) = 2.709, p < 0.05$). The group of Cert responses was significantly associated with the highest HR values suggesting that the more certain the students were about their answers, the more likely they tended to have positively valenced emotions. Positive valence for affective modeling -even if there is no wide agreement upon its interpretation- is more associated to positive emotions (Lang, 1995). This interpretation strengthens the intuitive fact that when a student is certain about his response, he tends to manifest a calm attitude and express positive emotions like satisfaction or joy. However, a state of uncertainty is usually related to negative emotions like confusion, dissatisfaction, or frustration.

No significant differences were found between the groups of answers for the GSR data ($F(3, 566) = 1.623, p = \text{n.s.}$), suggesting that the types of responses were not related to the intensity of the emotional reactions (arousal) but rather to their valence. In order to go further within this analysis, we considered the proportions of positive emotions within the target region of the Circumplex model, given by a slight positive valence and neutral arousal. We found a statistically significant difference of the mean of target emotional proportions across the four types of responses ($F(3, 566) = 3.361, p < 0.05$). We observed the highest proportions for the Cert responses, which suggests that the learners were more frequently within the target region when they were certain about their answers.

Based upon the obtained results, different trends in terms of concentration, valence, and positive emotional activations can be related to the state of uncertainty. Other interesting trends were observed looking at the response time across the four groups of responses; we found a statistically significant difference between the four conditions ($F(3, 566) = 137.925, p < 0.01$). The shortest response time were observed for the Cert responses ($M = 29.75, SD = 19.00$) compared to Low_Uncert ($M = 54.39, SD = 20.11$), High_Uncert ($M = 66.42, SD = 12.73$), and No_Resp ($M = 59.10, SD = 20.35$). Indeed, one can expect that a learner responds faster when he is certain about his answers than when he is uncertain and can take more time to try to figure out the solution. This natural tendency confirms that learners’ response time should be taken into account for an accurate uncertainty assessment.

In our next investigation we analyzed learners’ individual a priori tendencies to be in particular (un)certainty states. We examined the impact of personal characteristics namely age, personality traits, gender, and skill level in the logical based problem solving tasks, by looking at the eventual individual associations between these criteria and the number of answers for each level of (un)certainty.

First, we ran bivariate correlations to assess the relationships respectively between participants’ age and each of the five personality trait scales (OCEAN) and the number of answers for the different considered types of responses. No significant correlation was found with regards to the age variable. However, for the personality traits, statistically
significant Pearson’s correlation coefficients (r) were found for the conscientiousness trait scale. Table 1 summarizes the correlational results. Interestingly, we found a positive although low correlation between the conscientiousness trait scale and the number of Cert responses (r = 0.364), a low negative correlation with the number of Uncert responses (r = -0.399), and a moderate negative correlation with the number of High_Uncert responses (r = -0.501). These correlations were statistically significant (p < 0.05) suggesting that the more conscientious the participant was, the more he tended to be uncertain in the logical quiz.

Table 1. Bivariate correlational results

<table>
<thead>
<tr>
<th>Response Type</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cert</td>
<td>0.364*</td>
<td>0.025</td>
</tr>
<tr>
<td>Uncert</td>
<td>-0.399*</td>
<td>0.013</td>
</tr>
<tr>
<td>Low_Uncert</td>
<td>0.076</td>
<td>0.652</td>
</tr>
<tr>
<td>High_Uncert</td>
<td>-0.501**</td>
<td>0.001</td>
</tr>
<tr>
<td>No Resp</td>
<td>-0.019</td>
<td>0.908</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).

Second, one-way ANOVAs were performed to test the associations between the number of answers of each response type and respectively, participants’ perceived skill level in logical problem solving and the gender variables. A statistically significant effect of the skill level was found for the number of Cert responses (F(1, 36) = 4.346, p < 0.05) and for the number of High_Uncert responses (F(1, 36) = 4.268, p < 0.05). Results revealed that learners with moderate to high skill level had more certain answers (M = 9.18, SD = 2.23) than learners with low skill level (M = 7.56, SD = 2.52) and less highly uncertain answers (M = 2.18, SD = 1.65 versus M = 3.44, SD = 2.09). This suggests that participants with moderate to high skills in logical based problem solving were more certain about their answers in the logical quiz than participants with a low skill level. A significant effect of gender was also found for the number of High_Uncert responses (F(1, 36) = 4.872, p < 0.05). Women reported more highly uncertain responses (M = 3.57, SD = 2.13) than men (M = 2.21, SD = 1.64). These results underline the importance of individual characteristics/differences (such as gender, skill level, or personality traits) for a multidimensional modeling of uncertainty.

Uncertainty prediction

In the previous section we were interested in identifying indicators that can distinguish between learner trends and contribute in assessing levels of uncertainty. It was found that several facets from their electrophysiological activity as well as cognitive and personal parameters were significantly related to their state of uncertainty confirming hence our first hypothesis. In this section we are interested in the second hypothesis of this research that is a combination of these factors can reliably assess a user’s uncertainty level. We train classifier models by taking as an input, features that revealed statistically reliable associations with the (un)certainty levels namely delta_low_theta rate, HR, target emotions proportions, response time, gender, skill level and the conscientiousness trait scale.

First, we trained a binary classifier to predict the Uncert from the Cert responses. Then, we extended the analysis to predict users’ uncertainty in a more detailed level (High_Uncert, Low_Uncert, Cert). Besides, two separate datasets were considered. In the first dataset, No Resp samples were either included with the Uncert samples or gathered in a separate class. In the second dataset, No Resp samples were discarded. This separation is motivated by the ambiguous interpretation of the No Resp samples (10.17% of the data). Does a no-response mean a high level of uncertainty such that the learner was unable to reach the solution of the problem within the allowed time or did not take the risk to respond? Or does it merely indicate that the learner did not have the time to respond even if he knew the correct answer? Table 2 shows the accuracies of classification results from three machine learning algorithms namely Decision Tree (DT), Naïve Bayes (NB) classifier, and Support Vector Machines (SVM) (Witten & Frank, 2005).

Prediction performance was evaluated using a K-fold cross validation technique (Efron & Tibshirani, 1993). The input dataset is divided into K subsets. The classifier is trained on K-1 subsets and evaluated on the remaining subset. This process is repeated K times, the accuracy estimates are averaged to yield the overall classifier accuracy. This
study employed the Weka software (Witten & Frank, 2005), a collection of machine learning algorithms intended for data mining tasks. We used the software’s default parameters for the three algorithms with K = 20.

<table>
<thead>
<tr>
<th>Classes</th>
<th>DT</th>
<th>NB</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cert, Uncert</td>
<td>77.37%</td>
<td>78.42%</td>
<td>79.64%</td>
</tr>
<tr>
<td>Cert, Uncert, No Resp</td>
<td>72.46%</td>
<td>71.58%</td>
<td>73.33%</td>
</tr>
<tr>
<td>Cert, Low Uncert, High Uncert, No Resp</td>
<td>64.56%</td>
<td>63.68%</td>
<td>65.08%</td>
</tr>
</tbody>
</table>

Table 2. Classifier accuracy results | 1st dataset (No Resp included) |

<table>
<thead>
<tr>
<th>Classes</th>
<th>DT</th>
<th>NB</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cert, Uncert</td>
<td>78.90%</td>
<td>80.86%</td>
<td>83.25%</td>
</tr>
<tr>
<td>Cert, Low Uncert, High Uncert</td>
<td>73.84%</td>
<td>71.67%</td>
<td>74.46%</td>
</tr>
</tbody>
</table>

As presented in table 2, the SVM classifier has shown the highest prediction rates in all cases with accuracies ranging from 65.08% for the 4-class model (Cert, Low Uncert, High Uncert, No Resp) to 83.25% for the binary model (Cert, Uncert) excluding the no-answers from the training set (2nd dataset). Indeed, we noticed that merging the No Resp examples in the Uncert category slightly decreases the quality of the model to 79.64% (Cert, Uncert binary model in the 1st dataset), which suggests that trained models are clearly sensitive to the introduced inputs and hence that the no-answers can eventually involve both uncertainty and certainty states, which introduces a bias in the model.

<table>
<thead>
<tr>
<th>Classes</th>
<th>DT</th>
<th>NB</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cert, Uncert</td>
<td>69.66%</td>
<td>70.64%</td>
<td>72.07%</td>
</tr>
<tr>
<td>Cert, Uncert, No Resp</td>
<td>68.26%</td>
<td>67.22%</td>
<td>69.28%</td>
</tr>
<tr>
<td>Cert, Low Uncert, High Uncert, No Resp</td>
<td>57.80%</td>
<td>60.73%</td>
<td>62.08%</td>
</tr>
</tbody>
</table>

Table 3. Classifier accuracy results without the sensor data | 1st dataset (No Resp included) |

<table>
<thead>
<tr>
<th>Classes</th>
<th>DT</th>
<th>NB</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cert, Uncert</td>
<td>70.90%</td>
<td>72.07%</td>
<td>74.66%</td>
</tr>
<tr>
<td>Cert, Low Uncert, High Uncert</td>
<td>68.58%</td>
<td>67.32%</td>
<td>70.69%</td>
</tr>
</tbody>
</table>

These results confirm our second hypothesis. That is a classifier model of a user’s uncertainty state can be built on the basis of a multimodal combination of factors from affective variables, namely mental concentration, valence, and positive emotional activations, trait-questionnaire features such as the response time and individual differences such as gender, skill level, and personality trait scales. Results also suggest that this approach can be further extended to handle several levels of uncertainty. We believe that our method could be an appropriate alternative for an ITS to automatically assess users’ uncertainty states using machine learning techniques applied to EEG, GSR, and HR measures using non-intrusive sensors, as well as cognitive and personal criteria, so that ultimately, the prediction could be used to guide learning during computer-based education.

In order to highlight the contribution of these additional sensors, we replicated our analysis by excluding all the sensor data from the prediction models’ inputs. Table 3 shows the results of the classification results from the three algorithms using the same above setup. Accuracy rates ranged from 57.80% for the 4-class model (Cert, Low Uncert, High Uncert, No Resp) in the 1st dataset to 74.66% for the binary model (Cert, Uncert) excluding the no-answers in the 2nd dataset. Prediction performance decreased for the three algorithms in all the cases as compared to the previous approach including the sensor data. These results confirm that a multimodal sensor-based method yields more accurate predictive models. For instance, for the best-case binary model (Cert, Uncert), the prediction accuracy of the SVM classifier decreased from 83.25% to 74.66% for the same considered settings. This suggests that there is a non-negligible contribution and an obvious advantage of integrating affective data through these electrophysiological sensors to assess learners’ uncertainty states and that endowing ITS with capabilities to track learners’ mental and emotional reactions could give rise to more accurate user one-line monitoring and thereby eventually providing intelligent assistance and more efficient automated interventions and tutorial adjustments.
Conclusion and future works

In this paper we have proposed a new multimodal sensor-based approach to assess students’ uncertainty on the basis of their cerebral and emotional behavior using electrophysiological data with cognitive and personal variables. An experimental protocol was established by recruiting 38 participants to record EEG, BVP, and SC signals as well as trait-questionnaire responses and individual criteria namely age, gender, skill level, and personality trait scales. Participants interacted with a logical problem solving system designed to elicit uncertainty and reported their perceived level of uncertainty regarding each answer. These responses were used to supervise the classification process.

Results confirmed that students’ cerebral activity and emotional reactions in terms of mental concentration, HR, and positive target emotions with regards to the Circumplex model of emotions were significantly associated to different uncertainty levels. We also observed that participants’ individual differences contributed to some trends to be in particular uncertainty states. Finally, we developed classifiers to automatically predict levels of uncertainty using machine learning techniques, with the SVM algorithm demonstrating the best accuracy results (83.25%), and showed that a sensor-based modeling approach yields more precise predictions as opposed to a conventional modeling. This work should however be extended with a deeper comparative study with regards to current methods of uncertainty assessment.

Our future research trends will be focused on using this sensor-based approach to track the students’ states and guide the teaching process in a way that enhances users’ cognitive abilities and learning performance. In the short term, we are planning to extrapolate uncertainty models within more complex learning situations, gathering more data, and refining the models by incorporating further parameters from learners’ profiles. In the long term, we will be interested in developing a tutor that will integrate real time model predictions and select appropriate pedagogical strategies according to the classifiers’ outputs. The tutor will use associations between user’s actions and internal states to adjust the tutoring content. Further variables such as the frequency of user’s uncertainty, history of the presented concepts and system’s interactions, and answer correctness will be considered to track the subjacent potential cause of uncertainty and adapt the problem difficulty levels and the adequate support to the user.

Acknowledgments

We would like to acknowledge the National Science and Engineering Research Council (NSERC) and the Tunisian Ministry of Higher Education and Scientific Research for funding this work. We are also thankful to Maher Chaouachi for his valuable comments.

References


Graesser, A., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95*(3), 524-536


Exploring Factors Affecting Learner’s Perception of Learning Information and Communication Technology: A HLM Analysis of a National Farmers’ Training Program in Taiwan

Hsiu-Ping Yueh¹*, Tzy-Ling Chen², Li-An Chiu¹ and Wei-Chien Lin³

¹Department of Bio-Industry Communication and Development, National Taiwan University, Taiwan // ²Graduate Institute of Bio-Industry Management, National Chung Hsing University, Taiwan // ³ Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan // yueh@ntu.edu.tw // tlchen@dragon.nchu.edu.tw // melodyat1109@gmail.com // weichien.lin@gmail.com

*Corresponding author

(Submitted June 14, 2011; Revised October 26, 2011; Accepted November 15, 2011)

ABSTRACT
The present study, pertaining to a national information literacy training program for both farmers and rural communities in Taiwan, explores factors that affect learners’ perception of learning information and communication technology (ICT). It further analyzes effects of individual characteristics and varied training designs on learners’ perception of learning ICT. Data used for analyses derive from the Farmers’ ICT Training Project of 2005, conducted by the Council of Agriculture in Taiwan (COAT), including evaluations of 4,405 trainees. Findings of hierarchical linear model (HLM) analysis supported both predictors of personal and organizational levels. Results of this study also revealed that learners’ characteristics and varied training designs do have effects on learner’s perception of training effectiveness. Specific impacts of learners’ personal characteristics on their perception of learning effectiveness in different types of ICT training are also analyzed and discussed. In conclusion, implications founded on learner’s perspective and suggestions for future research are illustrated in this paper.

Keywords
Learner’s perception, Information and communication technology (ICT) training, Perception of training effectiveness, Hierarchical linear model (HLM)

Introduction
It is suggested that information and communication technology (ICT) can offer disadvantaged communities such as farmers, long recognized as mute victims of such “clusters of disadvantage”, new solutions to a lack of developmental opportunities (Amirtham & Joseph, 2011). To promote application of ICT, providing appropriate training for practitioners is considered to be one of the most critical factors in their effective use of ICT (Holmes, 2009). In Taiwan, to help farmers use the Internet and ICT to improve their agricultural work and strengthen their competitiveness in agribusiness management, the Council of Agriculture in Taiwan (COAT) initiated a national Farmers’ ICT Training program in 2002. Although the primary trainees are farmers, rural inhabitants are included as secondary subjects. Non-farming rural inhabitants are included mainly because one of the purposes of the training initiative is to enhance the ICT literacy of rural communities in Taiwan. In addition, these inhabitants are an important human resource available to farmers when they need extra hands or help during busy seasons. It is hoped that this program can increase the ICT literacy of both farmers and non-farming rural inhabitants. The primary focus of this study is the ICT training of farmers and other rural inhabitants in Taiwan, and the study employs data from the Farmers’ ICT Training Database, a national ICT training survey.

Understanding what factors contribute to learner’s perception of training effectiveness can help program planners develop an effective, useful, and beneficial training program. This paper describes the relationship among personal demographic characteristics of learners, organizational characteristics, and trainees’ perception of learning ICT. Four criteria are used to measure trainees’ perception of learning or training effectiveness of the ICT training program, including the perceived satisfaction, perceived usefulness, mastery of learning, and confidence of learning transfer.

In technology-related training literature, research on the influence of individual differences has always been a popular topic. Bartel and Lichtenberg (1987) revealed that highly educated workers have a comparative advantage with respect to their adjustment to and utilization of new technologies. Since specific experiences with technology
are considered important for adoption and implementation of new types of technology, educated trainees can more easily learn to use a microcomputer and its accompanying software (Huffman & Mercier, 1991). Mathieu, Tannenbaum, and Salas (1992) proved the influence of education level on learning based on a model that takes into consideration the individual and situational influences on learning motivation and learning effectiveness. Morris (1994) also confirmed that personal attributes of participants can result in a positive influence on their learning in a computer course. An empirical research on extension programs conducted in Australia in 1996, which explored environmental concern and actions of farmers, indicated that farmwomen were better educated and more likely to be environmentally concerned; thus, it is suggested the related educational programs and training should be targeted at women (Geno, 2002). In summary, many studies have focused on individual differences in training research, and personal traits such as age, educational background, experience with specific software, overall computer experience, gender, and years of education are often considered (Bostrom, Olfman, & Sein, 1990; Sun, Tsai, Finger, Chen, & Yeh, 2008).

Prior studies also revealed the importance of a research focus on individual difference variables that are associated with learning about end-user software. Lee (1997) reported that computer or programming experts may have great technical skills but not be aware of certain aspects of adult learning characteristics when planning related training programs. If training fails to help adults learn, it is ineffective. Adults who attend such types of ICT training thus remain anxious or unskillful in using ICT. Adult learning programs and courses need to be organized based on the learners’ expectations, profiles, and needs (Jimoyiannis & Gravani, 2011). Therefore, individual differences among potential trainees should be taken into account when developing a training program (Chou, 2001; Jimoyiannis & Gravani, 2011; Lim, Lee, & Nam, 2007; Liu & Reed, 1994). In this study, the variables of age, gender, education level, occupation, and ICT-related experience of learners are incorporated for investigation.

According to Noe (1986), who integrated motivational and situational factors derived from organizational behavior theory into the research model, trainees’ attributes and attitudes towards training influence their perceived effectiveness of training. Also, a well-organized and -managed training process is beneficial to older adults’ learning of new computing technologies (Mantzana, Themistocleous, & Morabito, 2010). Baldwin and Ford (1988), Aguinis and Kraiger (2009), and Bates (2004), on the other hand, confirmed that both trainees’ personal characteristics and organizational environments affect learning performance of training. However, few prior studies have related organizational characteristics to individual learning outcomes in training. Although there is a lack of sufficient research endeavors, based on institutional theory (Orlikowski, 1992), human behaviors are recognized under the influences of structural and institutionalized properties, called “structure”, where they happen and exist. Since the interaction between the organizational environment where training takes place and training effectiveness should be considered equally important in training research (Lim, Lee, & Nam, 2007), another aim of the present study is to address a such deficit. To accomplish this task, a hierarchical linear model (HLM) was employed to verify relationships among learner characteristics (gender, age, education level, occupation, and ICT-related experience), organizational characteristics (level of training, type of training host, and region of training delivery), and learner’s perceived training effectiveness (learner’s perceived satisfaction, perceived usefulness, mastery of learning, and confidence of learning transfer).

HLM is an analysis method frequently used in many research fields, such as education (Clapp et al., 2007; Fleming, Harachi, Catalano, Haggerty, & Abbott, 2001; Kutash, Banks, Duchnowski, & Lynn, 2007; Ma & Wilkins 2002; Spencer, Noll, & Cassidy, 2005), social policy (Feinberg, Greenberg, & Osgood, 2004; Jasuja et al., 2005), and public health (Hser, Joshi, Maglione, Chou, & Anglin, 2001; Johnson, Young, Suresh, & Berbaum, 2002; Schutt, Rosenheck, Penk, Drebing, & Seibyl, 2005; Woodward, Das, Raskin, & Morgan-Lopez, 2006). In the cases of education, social policy, and public health, participants under research belong to a certain specific context or structure, such as a school or class, social system, or program, respectively. Thus, they will be influenced by some structural or contextual factors consisting of, but not limited to, the interaction norms, support, resources, and behavioral regulations. To determine the individual effect and interplay of influences of a participant’s personal level and organizational level, HLM can deal with different kinds of influences and perform a complete analysis on data with varied sources of influence (Raudenbush & Bryk, 2002).

In a study by Woodward, Das, Raskin, and Morgan-Lopez (2006), the HLM method was used to address the interaction between organizational characteristics and treatment outcomes of alcohol and drug services in the United States. The results revealed that the patient’s demographics, such as age, education level, employment status, treatment duration, and previous substance abuse treatment, as well as organizational factors, such as the education level, occupation, and organizational level, HLM can deal with different kinds of influences and perform a complete analysis on data with varied sources of influence (Raudenbush & Bryk, 2002).
level of faculty members, counseling personnel costs, and treatment type, are significantly associated with treatment completion. In addition, an interaction was observed between organizational type and number of intake diagnoses by patients. As for training evaluation, both Feinberg, Greenberg, Osgood, Anderson, and Babinski (2002) and Lundgren and Rankin (1998) applied HLM to their studies. The findings by Feinberg et al. (2002) presented that training is positively linked with the attitudes and knowledge of community leader participants, but is not related to their perceived community readiness or perceived efficacy of communities that care. Moreover, it is evidenced that long-term influences of training in this context may be found at the group or coalition level, but not at the individual level. Lundgren and Rankin (1998) found that both client and program characteristics affect employment wages. The influence of type of training is particularly critical, and for those young graduates, the wage received after training is further associated with client characteristics.

**Purpose of the study**

Considering that both personal and organizational effects on learner’s perception of training effectiveness are important, it is necessary to expand research concerns not only to incorporation of these two determinants but also to analysis of their interplay. Despite the potential significance of the impacts, a review of the literature indicates a lack of related research so far. To bridge this gap in research, the main purpose of this study is to take into account and analyze what personal and organizational factors are linked to learner’s perception of the effectiveness of Farmers’ ICT training, while simultaneously exploring how these two sets of predictors are related to each other using the HLM method.

**Research method**

**Farmers’ ICT training database**

The Farmers’ ICT Training Database was developed with the support of COAT, which includes data related to training evaluation conducted at the end of each training program since 2002. The Farmers’ ICT Training Database collects information about learners’ demographic characteristics and background data, such as years of farming practice, types of crops planted, and farm size. Also, at the end of each session, learners are required to complete a post-training questionnaire online, and the survey data are then included in this database. This questionnaire collects learner’s perceptions of training satisfaction, usefulness of training materials, mastery level of ICT use, and confidence in applying what they have learned in the training to their practice or daily lives. By taking advantage of this database, this study attempts to promote a better understanding of the relationship among trainee’s characteristics, organizational features of training, and training effectiveness as measured by learner’s perceived learning effectiveness, one of the most important aspects of interpreting training effectiveness.

**Data sample and research question**

The study used survey data of 2,671 learners who completed Farmers’ ICT Training provided by COAT. Of these, 1,291 trainees were participants in basic classes, and 1,380 trainees, in advanced classes. The participants in this study were 46 years old on average; 71.3 percent of them were female; 54.6 percent had a senior high school education; and all had common ICT experiences. In addition, 58.4 percent of participants were non-farmers. High-school educated females aged 40-49 years were the major participants in both basic and advanced training classes. Non-farmer participants outnumbered farmer participants in both classes. The participants in both classes had some experience with ICT, but participants who attended advanced training were more familiar with ICT. All participants had positive perceptions of learning new technology.

In addition to personal characteristics, this study also used data of organizational characteristics and learner’s perception of training effectiveness obtained from the Farmers’ ICT Training Database in order to answer the research questions. Drawing upon the purposes of the study, three research questions are listed as follows:

- Is there any relationship between trainees’ personal characteristics and perception of learning ICT?
- Is there any relationship between training-related organizational characteristics and trainees’ perception of learning ICT?
• Do training-related organizational characteristics moderate the relationship between trainees’ personal characteristics and perception of learning ICT?

Research framework and variable

The research framework, shown in Figure 1, was developed in line with the questions for analysis. Two datasets of the Farmers’ ICT Training Database were utilized in this study. One consisted of personal demographic data for 4,405 trainees, and the other included survey data related to learner’s effectiveness perception of learning ICTs in Farmers’ ICT Training, as measured by perceived training satisfaction, usefulness, mastery level of ICT use, and confidence in learning transfer. For HLM analyses, learner’s perception data were utilized as the dependent variables. Used as independent variables were five personal characteristics of trainees: gender, age, education level, occupation, and ICT-related experience of learners.

As illustrated in the research framework, three training-related organizational characteristics were also used as independent variables in this study. The first one concerns level of training (basic vs. advanced). Basic ICT training focuses on an introduction to computers and the Internet, and basic application of e-mail and website browsing. Advanced ICT training focuses on application of the Microsoft Office software package, such as Word, PowerPoint, and so on, and other related commercial application software for web publishing, posting, and advertisement production. Variation of training design distinguishes between the different learning task requirements of trainees and is assumed to cause differences in the learner’s effectiveness perception of learning ICT.

The second training-related organizational variable is training host. This variable is defined according to the administrative level of training providers, as well as how the training program is implemented. At both the county/city and local township levels, some organizations host the training by themselves, and others outsource the training to other organizations. In this case, three types of training hosts in 2005 can be classified: local self-provider, self-provider of county/city level, and outsourcing-host of county/city level.

The third one is related to region of training delivery. Since the resource distribution and overall economic conditions are not equivalent throughout Taiwan, regional development varies to some degree. As a result, the developmental features and resources related to ICT of different counties and local areas are varied. The factor of region of training delivery must be considered to examine whether variance in learner’s perception of training effectiveness exists among different regions. The region of training delivery is classified into the northern, central, and southern sections of Taiwan.

---

Figure 1. Research Framework
In the current study, the numbers of advanced and basic training classes offered were about the same. However, a higher percentage of training was self-hosted by the local farmers’ associations (54.4%) than by outsourced-hosts and providers of the county/city level. Also, a higher percentage of training was delivered in the central region of Taiwan (49.1%). The descriptive analysis of training-related organizational variables is showed in Table 1 below.

<table>
<thead>
<tr>
<th>Table 1. Descriptive analysis of training-related organizational variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizational Variables</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Level of training</td>
</tr>
<tr>
<td>Advanced</td>
</tr>
<tr>
<td>Basic</td>
</tr>
<tr>
<td>Type of training host</td>
</tr>
<tr>
<td>Local self-provider</td>
</tr>
<tr>
<td>Outsourcing-host of county/city level</td>
</tr>
<tr>
<td>Self-provider of county/city level</td>
</tr>
<tr>
<td>Region of training delivery</td>
</tr>
<tr>
<td>Central region</td>
</tr>
<tr>
<td>Southern region</td>
</tr>
<tr>
<td>Northern region</td>
</tr>
<tr>
<td>Total classes</td>
</tr>
</tbody>
</table>

As for learner’s perceptions of training effectiveness, four dependent variables were employed to explore trainees’ perception of learning ICT. The first was perceived satisfaction, which was constructed to understand how satisfied learners felt about the ICT training held by COAT. The perceived satisfaction of training contained 11 statements to evaluate learner’s perception of the instructors, materials used, and learning process and environment of training. Participants were asked to indicate their degree of agreement on a six-point Likert-type scale ranging from strongly agree (5) to strongly disagree (0). The perceived usefulness measures how useful the participants see the training as being, as regards materials used and learning contents offered. The perceived usefulness can help designers or program planners to understand whether the current training is useful and appropriate or not. The third criterion was perceived mastery of learning, which means the perceived personal learning outcomes. The fourth criterion was perceived confidence of learning transfer, which means learner’s confidence of applying what they have learned in the training to their practice or daily lives. Due to differences in training design, the measures of the three criteria of perceived usefulness, mastery of learning, and confidence of learning transfer consisted of 13 question items for the basic training and 8 items for the advanced training. Participants were asked to indicate their degree of agreement on a three-point Likert-type scale ranging from strongly agree (3) to strongly disagree (1). Higher scores on the measures indicated higher degrees of learner agreement on the perceived usefulness, learning mastery, and confidence in applying what they have learned to their future practice or daily lives.

**Assessment of perception measurement**

To ensure the validity of the measurement of learner’s perception, the construction of the post-training questionnaire first followed a systematic review of related literature in combination with an evaluation of the objectives of Farmers’ ICT Training. When completed, the questionnaire was further validated by experts in the field of agricultural extension and ICT training professionals to obtain expert validity. To test the reliability of measurement of learner’s perceived training effectiveness used in this study, Cronbach’s α was adopted to represent internal consistency. Table 2 shows the results of the reliability test, with all values greater than the threshold of 0.70. The results show that the measurement reliably measures the defined constructs and variables.

<table>
<thead>
<tr>
<th>Table 2. Reliability and descriptive analysis of learner’s perceived training effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic class</td>
</tr>
<tr>
<td>Cronbach’s α</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>perceived satisfaction</td>
</tr>
<tr>
<td>perceived usefulness</td>
</tr>
<tr>
<td>perceived mastery level of learning</td>
</tr>
<tr>
<td>perceived confidence of learning transfer</td>
</tr>
</tbody>
</table>
Results

In this section, the correlations between the variables targeted for research are presented (Table 3). Subsequently, the influences of learner’s personal- and organizational-level predictors on the perceived training effectiveness are addressed. Finally, the relationships among learner’s personal characteristics, training-related organizational characteristics, and learner’s perception of training effectiveness are explored by means of HLM, and the results reveal the confirmed moderating effects of one organizational-level predictor.

Analysis of correlations

Table 3. Correlation analysis of research variables

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>--</td>
<td>--</td>
<td>1</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>Age</td>
<td>46.64</td>
<td>9.82</td>
<td>.12**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education Level</td>
<td>--</td>
<td>--</td>
<td>.04*</td>
<td>.26**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>--</td>
<td>--</td>
<td>.45**</td>
<td>.01</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience with ICT</td>
<td>3.65</td>
<td>0.50</td>
<td>.01</td>
<td>-.21**</td>
<td>.27**</td>
<td>-.06**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived satisfaction</td>
<td>4.62</td>
<td>0.52</td>
<td>-.03</td>
<td>.07**</td>
<td>-.03</td>
<td>-.05*</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>2.53</td>
<td>0.48</td>
<td>.03</td>
<td>-.04</td>
<td>.06**</td>
<td>-.04**</td>
<td>.03</td>
<td>.29**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived mastery level of learning</td>
<td>1.9</td>
<td>0.53</td>
<td>.05**</td>
<td>-.14**</td>
<td>.15**</td>
<td>-.01</td>
<td>.20**</td>
<td>.20**</td>
<td>.54**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived confidence of learning transfer</td>
<td>2.09</td>
<td>0.53</td>
<td>.08**</td>
<td>-.08**</td>
<td>.14**</td>
<td>.02</td>
<td>.16**</td>
<td>.22**</td>
<td>.57**</td>
<td>.74**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Level of training</td>
<td>1.49</td>
<td>0.62</td>
<td>.04</td>
<td>-.11**</td>
<td>.05**</td>
<td>10**</td>
<td>.02</td>
<td>.05*</td>
<td>-.10**</td>
<td>.01</td>
<td>.00</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Training boxtypes</td>
<td>0.52</td>
<td>0.50</td>
<td>-.03</td>
<td>-.13**</td>
<td>.19**</td>
<td>-.04</td>
<td>.21**</td>
<td>-.09**</td>
<td>-.12**</td>
<td>.01</td>
<td>-.04*</td>
<td>.03</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2 Training delivery region</td>
<td>1.2</td>
<td>0.68</td>
<td>.10**</td>
<td>-.05*</td>
<td>.03</td>
<td>10**</td>
<td>-.02</td>
<td>.02</td>
<td>-.02</td>
<td>.05**</td>
<td>.07**</td>
<td>.41**</td>
<td>.01</td>
<td>1</td>
</tr>
</tbody>
</table>

*p<.05; **p<.01

According to McMillan (2000), “correlation coefficients between 0.10 and 0.30 are referred to as having small or weak positive relationships, between 0.40 and 0.60 as moderate positive relationships, and 0.70 and above as high positive relationships.” As presented in Table 3, the correlation coefficients among perceived satisfaction, perceived usefulness, mastery level of learning, and perceived confidence of learning transfer are examined. The most important figure is the high significant correlation between learner’s perceived confidence of learning transfer and mastery level of learning ($\gamma=.739$, $p<.01$). Also, significantly moderate correlations between learner’s perceived usefulness and perceived confidence of learning transfer ($\gamma=.567$, $p<.01$) and perceived mastery level of learning ($\gamma=.538$, $p<.01$) are confirmed. This indicates that with greater belief in their own capabilities to apply the skills learned, trainees are more likely to think they will make use of these skills in the future. Similarly, more prominent interactions among trainees’ perceived usefulness, mastery level of learning, and confidence of learning transfer are further evidenced, although some correlation relationships among the other research variables are found to be significant, too.

Learner’s personal-level predictors

Based on the analysis results of HLM shown in Table 4, it is clear that all learners’ personal characteristics have influences on some of their effectiveness perception of learning ICT. The variables of gender and education level were proved to have influences on perceived usefulness, mastery level of learning, and confidence of learning transfer. In particular, three perception scores of male participants were higher than those of female participants. The learners with better education also had more positive perceptions in these three training effectiveness variables. Learner’s occupation had influences only on the perceived usefulness. The non-farmer learners had more positive perception of training usefulness than did farmer participants.

In terms of learner’s age, it was revealed that age has influences on four variables relevant to training effectiveness. The older trainees tend to have more positive perceptions on perceived satisfaction, but less on perceived usefulness, mastery level of learning, and confidence of learning transfer. Learner’s ICT-related experiences also affect their perceived satisfaction, mastery level of learning, and confidence of learning transfer. Learners with more ICT experiences will tend to be more familiar with ICT and to have more positive perceptions.
### Table 4. Random-coefficients-regression model of training effectiveness

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Perceived satisfaction</th>
<th>Perceived usefulness</th>
<th>Perceived mastery level of learning</th>
<th>Perceived confidence of learning transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect</td>
<td>( \gamma )</td>
<td>( t )</td>
<td>( \gamma )</td>
<td>( t )</td>
</tr>
<tr>
<td>Intercept ( \gamma_{00} )</td>
<td>4.255</td>
<td>37.264***</td>
<td>2.533</td>
<td>14.279**</td>
</tr>
<tr>
<td>Gender ( \gamma_{10} )</td>
<td>-0.039</td>
<td>-1.484</td>
<td>0.058</td>
<td>2.602**</td>
</tr>
<tr>
<td>Age ( \gamma_{20} )</td>
<td>0.004</td>
<td>2.741*</td>
<td>-0.002</td>
<td>-2.044*</td>
</tr>
<tr>
<td>Education Level 1 ( \gamma_{30} )</td>
<td>-0.001</td>
<td>-0.028</td>
<td>0.052</td>
<td>2.401†</td>
</tr>
<tr>
<td>Education Level 2 ( \gamma_{40} )</td>
<td>-0.039</td>
<td>-1.219</td>
<td>0.078</td>
<td>2.567*</td>
</tr>
<tr>
<td>Occupation ( \gamma_{50} )</td>
<td>-0.029</td>
<td>-1.076</td>
<td>-0.044</td>
<td>-2.238†</td>
</tr>
<tr>
<td>Experience with ICT ( \gamma_{60} )</td>
<td>0.059</td>
<td>2.540*</td>
<td>0.015</td>
<td>0.857†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance components</th>
<th>( \tau )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 residual ( \gamma )</td>
<td>0.211</td>
<td>0.143</td>
</tr>
<tr>
<td>Intercept ( \tau_{00} )</td>
<td>0.415</td>
<td>112.352**</td>
</tr>
</tbody>
</table>

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

### Training related organizational-level predictors

Similarly, the HLM analysis results of Table 5 support that level of training influences perceived satisfaction and usefulness. The trainees in basic classes perceived higher degrees of satisfaction and usefulness than did those in the advanced training. This result may indicate that it is much more difficult to determine what learning tasks are suitable for inclusion in the classes of advanced levels, since participants’ entry knowledge or prior abilities with respect to ICT can vary more greatly than those of trainees entering basic or entry-level classes. In this investigation, the result was a lower perceived training effectiveness of learners in the advanced ICT training. The characteristics of training host and region of training delivery did not lead to any significant influences on effectiveness perception of learning ICT, according to the analysis results.

### Table 5. Intercepts-as-outcomes model of training effectiveness

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Perceived satisfaction</th>
<th>Perceived usefulness</th>
<th>Perceived mastery level of learning</th>
<th>Perceived confidence of learning transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect</td>
<td>( \gamma )</td>
<td>( t )</td>
<td>( \gamma )</td>
<td>( t )</td>
</tr>
<tr>
<td>Intercept ( \gamma_{00} )</td>
<td>4.224</td>
<td>32.867***</td>
<td>2.717</td>
<td>23.907***</td>
</tr>
<tr>
<td>Level of training ( \gamma_{01} )</td>
<td>-0.091</td>
<td>-2.661**</td>
<td>-0.136</td>
<td>-3.133***</td>
</tr>
<tr>
<td>Training host 1 ( \gamma_{02} )</td>
<td>0.102</td>
<td>1.408</td>
<td>-0.062</td>
<td>-0.682†</td>
</tr>
<tr>
<td>Training host 2 ( \gamma_{03} )</td>
<td>0.016</td>
<td>0.203</td>
<td>-0.124</td>
<td>-1.230†</td>
</tr>
<tr>
<td>Delivery 1 ( \gamma_{04} )</td>
<td>0.004</td>
<td>0.074</td>
<td>-0.048</td>
<td>-0.628†</td>
</tr>
<tr>
<td>Delivery 2 ( \gamma_{05} )</td>
<td>0.047</td>
<td>0.783</td>
<td>-0.017</td>
<td>-0.218†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance components</th>
<th>( \tau )</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 residual ( \gamma )</td>
<td>0.211</td>
<td>0.142</td>
</tr>
<tr>
<td>Intercept ( \tau_{00} )</td>
<td>0.378</td>
<td>110.886***</td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001
Interactions between personal- and organizational-level predictors

The HLM analysis of this study used the slopes-as-outcomes model to explore interactions between learner’s personal- and organizational-level predictors. This model also tested the moderating effect of organizational variables. As shown by the results in Table 6, the possibility of a moderating effect of organizational variables may exist in two training effectiveness variables, which are perceived satisfaction and perceived confidence of learning transfer. The results of analysis shown in Table 6 reveal that region of training delivery moderated the relationship of learners’ experiences with ICT in relation to their perceived satisfaction. Also, region of training delivery moderated the relationship between age of learners and their perceived confidence of learning transfer. Therefore, empirical evidence supporting the interaction term has been found.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>percieved satisfaction</th>
<th>perceived confidence of learning transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Variable</td>
<td>Gender</td>
<td>Age</td>
</tr>
<tr>
<td>Level of training</td>
<td>-0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>Training host types 1</td>
<td>0.011</td>
<td>0.005</td>
</tr>
<tr>
<td>Training host types 2</td>
<td>-0.006</td>
<td>0.007</td>
</tr>
<tr>
<td>Training delivery 1</td>
<td>0.062</td>
<td>0.005</td>
</tr>
<tr>
<td>Training delivery 2</td>
<td>0.050</td>
<td>0.005</td>
</tr>
</tbody>
</table>

*p<0.05; **p<0.01

Discussion and conclusion

Findings and discussion

It is worthwhile and beneficial to program planners to improve learning effectiveness and transfer performance by understanding better what affects how learners view and value training. In this study, the effect of learner’s personal-level predictors is confirmed, based on HLM analysis, to be associated with the perceived training effectiveness. The key factors of gender, age, and education level of learners are further supported by this research. The findings of the present study on the effects of learners’ personal characteristics on their perception of ICT training effectiveness are, overall, consistent with the literature (Bostrom, Olfman, & Sein, 1990; Chou, 2001; Liu & Reed, 1994; Mathieu, Tannenbaum, & Salas, 1992; Morris, 1994).

In particular, learner’s age has influences on all training effectiveness criteria used in the present study, including perceived satisfaction, usefulness, mastery level of learning, and confidence of learning transfer. This corroborates the study results of Gist et al. (1988), Lee (1997), and Mayhorn, Stronge, McLaughlin, and Rogers (2004). Similarly, gender and education level of learners have influences on three criteria, namely the perceived usefulness, mastery level of learning, and confidence of learning transfer. This result is also in accord with research findings by Chou (2001), Davis and Davis (1990), and Whitley (1997). Moreover, a higher education level is related to better perception of learning effectiveness. Learner’s occupation influences only the perceived usefulness. This indicates that to most adult learners, the ultimate goal of learning in a training program is to make good use of what is learnt afterward in the workplace. In addition, learners’ ICT experiences are related to the perceived satisfaction, mastery level of learning, and confidence of learning transfer. As for participant’s occupation, this study had two categories, farmer and non-farmer. Results showed that non-farmer participants had a higher perceived usefulness than farmer participants. For the most part, those non-farmer participants were farmers’ wives, who typically help take care of bookkeeping, accounting, and some routine business of farm management. They thus considered the ICT training to be more beneficial to their record keeping in terms of the perceived usefulness than did the farmer participants.

Not only is the effect of learner’s personal-level predictors confirmed, but also the influences of training-related organizational-level predictors on the perceived training effectiveness of learners are supported in this study. According to the results of analyses, the level of training has influences on learner’s perceived satisfaction and usefulness. Learners who attended the basic ICT training demonstrated a higher score of perceived satisfaction and usefulness than those who attended the advanced classes, which is consistent with Lim and Morris (2006). The variance in complexity of learning tasks designed for these two training levels led to a decrease in learner’s...
perceived satisfaction and usefulness in advanced training. On the other hand, the factors of training host and region of training delivery were found to have no effect at all, possibly because a set of standardized training materials was used and pre-training orientation was implemented for the trainers or instructors recruited for this program. In consequence, learners sensed less or no distinction regarding learning contents, training materials, instructional methods, and so on during training, even though training was given by different training hosts and learning occurred in different geographic locations.

Moderating influences were found between the learner’s personal- and organizational-level predictors. The results revealed that both trainees’ perceived satisfaction and confidence of learning transfer are moderated to some extent. Specifically, the relationship between learner’s ICT-related experience and perceived satisfaction was moderated by the region of training delivery, and interaction between learner’s age and perceived confidence of learning transfer was also moderated by the region of training delivery. In other words, when learner’s perceived training effectiveness is under the influence of some personal characteristics, such as ICT-related experience and age, the regional context where training is delivered plays a certain role in adjusting their interactions critically. This finding implies a need for additional study to determine the particular contextual influences on the learning effectiveness of ICT training developed for rural communities.

Implications and conclusion

One of the most unique contributions of this study is that it helps illuminate the interactions affecting the perceived training effectiveness based on learner’s standpoint, training-related organizational structure, and learner’s personal profile as a whole. This corroborates what is suggested by White, Altschuld and Lee (2008): Utilizing a comprehensive approach to plot for the evaluation of any kind is pivotal because there are so many possible variables and factors worthy of consideration. An evaluation like this one has not been systematically performed in the past because of a lack of a well-rounded database for such analysis. The availability of a national Farmers’ ICT Training Database makes the present research possible. Along with the continuation of similar farmers’ ICT training programs and with the assistance of this database, further research should consider enhancing the evaluation model to include new variables, such as the time used for training in different ICT subjects, objective training, or learning effectiveness measures or instructors’ viewpoints, by designing new surveys and including more contextual data in the analysis.

Moreover, it is found that a specific focus on evaluation of the farmer learner’s perception of ICT training effectiveness will be necessary in the future. Even though the current study attempted to further examine whether farmer-related factors would result in differences in evaluation, little was found. Based on the limited information of the sample of 1,110 farmer participants in this study, results of descriptive analyses showed that participants with larger farms and more years of farming practice scored better on the perceived satisfaction. Participants with more years of farming practice also illustrated a lower perception in mastery level of learning. However, additional comparative analyses of variance revealed no significant statistical difference in the two abovementioned variables of training effectiveness. No difference in training effectiveness was evidenced for either farm size or the communities like farmers’ associations or agricultural production and marketing groups the participants belong to. Since complete farmer-related personal characteristics were not obtained in the present study, it is impossible to probe deeper into their influences on farmer trainees. Novelty and extra value can be added to such research by analyzing more business related characteristics, such as types of agriculture and area of the farm, pertaining to this homogenous group of farmer learners.

Research to explore factors affecting learner’s perceived training effectiveness is not only beneficial to the design and planning of better training programs but also contributes to a more learner-centered instructional design of training. The importance of learner’s personal characteristics in the current study shows that concern for the ICT or the digital divide in rural communities is still necessary. Older trainees and those with less education were found to have less ICT-related experience, while the older participants also demonstrated a better effectiveness perception of learning ICT. It is suggested that when such participants are trained in some ICT skills, support such as public access to computer facilities and Internet connections should be offered to help them keep practicing the skills learned. This consequently would promote the ICT capabilities of trainees of rural communities, as expected by COAT, which initiates and provides such training services.
Despite a few research limitations, such as a lack of objective measures of training effectiveness and the perspectives of training hosts or instructors, and the limited data sample of trainees from rural communities, the current study contributes to the existing research and practices in the field of ICT training. The analyses allowed not only validation of the effects of the learner’s personal characteristics on training performance but also justification of some training-related organizational features of ICT training program planning. As to a more learner-oriented or learning-centered approach, in the context of ICT training designed for rural communities, the emphasis of program planning is better placed on hands-on experiential learning rather than on teaching skills. Consequently, participants will develop the ability to acquire, process, store, retrieve, and use information by taking advantage of “technologies or tools” practiced, which is considered essential to successful learning of ICT. To ensure the provision of high-quality training services, it is necessary to consider contextual influences to put a real learner-centric paradigm into practice. According to the aforementioned results, one exciting implication is the support of empowering some non-farmer learners to become active facilitators of promoting and supporting future ICT learning in this context. In essence, a more self-directed or informal learning approach is suggested, in combination with training of non-farmer participants, when re-designing the farmers’ ICT training program. Gradually, a flexible system of rural ICT education and training can be developed by degrees to meet more diverse lifelong learning needs.

Acknowledgements
The authors would like to thank the Council of Agriculture in Taiwan (COAT), National Science Council (NSC95-2520-S-002-001) and National Taiwan University for their grants in support of this study.

Reference


Lundgren, L., & Rankin, B. (1998). What matters more: The job training program or the background of the participant? An HLM analysis of the influence of program and client characteristics on the wages of inner-city youth who have completed JTPA job training. Evaluation and Program Planning, 21(1), 111–120.


Graphical Interaction Analysis Impact on Groups Collaborating through Blogs

Georgios Fessakis, Angelique Dimitracopoulou and Aggelos Palaiodimos*
Learning Technology & Educational Engineering Laboratory University of the Aegean, Demokratias 1, 85100
Rhodes, Greece // gfesakis@rhodes.aegean.gr // adimitr@aegean.gr // psemms07036@rhodes.aegean.gr
*Corresponding author

(Submitted June 19, 2011; Revised November 29, 2011; Accepted December 27, 2011)

ABSTRACT
This paper presents empirical research results regarding the impact of Interaction Analysis (IA) graphs on groups of students collaborating through online blogging according to a “learning by design” scenario. The IA graphs used are of two categories; the first category summarizes quantitatively the activity of the users for each blog, hence permitting the comparison of the students’ activity level in a group, while the second category allows the comparison among different groups. The statistical analysis of the students’ interactions shows the significant impact of the graphs presence based on the number of posts and comments produced by the groups. Furthermore, the graphs of the first category (intra-group IA) have stronger impact than the graphs of the second (intergroup IA). The results support the general claim that interaction analysis is an important component of self-regulation in computer supported collaborative learning environments. In addition the research conclusion suggests that collaborative learning by design is an effective model especially when combined with interaction analysis tools.

Keywords
Blogs, Learning by design, Interaction analysis, CSCL, E-Learning, Higher education

Introduction
Increasing interest has recently emerged on the utilization of social software (Allen, 2004) by the educational community. This occurs not only due to the increased availability of social applications on the internet (e.g., web 2.0 services) but also because of their consistency to modern learning theories. Indeed, the use of social software for educational purposes is directly related to the sociocultural theory (Vygotsky, 1986) and to the social constructivism theory (Ernest, 1994; Kim 2001). These theories advocate the importance of the learners’ interaction during active participation to learning activities which provide opportunities for design and construction of meaningful artefacts. Blogs (“Blog,” n.d.) constitute special instance of social software that is a network application supporting groups of actors in communication and interaction. In a general overview of educational uses of blogs, Downes (2004) notices that students participating in blogging have opportunities to (a) reflect on their texts; (b) engage in writing for significant time intervals; and (c) trigger long dialogue with their readers leading to new writing cycles. In teaching and learning, the blogs can be used (indicatively) to (“Blog,” n.d.): collect learning resources and share ideas and experiences; log notes and observations during an inquiry learning activity; manage a project; develop dialogue like in an online forum; reflect and communicate with teachers and peers-students; develop collaboration and social skills; obtain the motivation of writing for readers who comment you in order to participate more actively in the course; run online school newspapers, etc.

In Chen et al. (2005) the researchers integrated blogging with the learning portfolio approach highlighting the importance of the adoption of a well-defined pedagogical approach for the successful integration of blogs or any other content management model (e.g., wikis). The study of the impact of the kind of work or the genre of learning activity (e.g. project, problem solving, brainstorming, etc.) to blogging used for educational purposes appears as interesting field of research. In this direction Fesakis et al. (2008) studied the combination of blogs to learning by design pedagogical approach (Han & Bhattacharya, 2001) and claimed that teachers can utilize blogs in order to increase the communication and interaction among the students as well as enhance their participation and active engagement in the classroom. The research evidence showed that blogs fulfill the requirements of the learning approach in an acceptable level. Moreover, students claimed that the blog support for the learning by design activity communication and information management was decisive. Through the use of blogging students had increased opportunities to receive feedback though comments of both fellow-students and teachers. The blog assisted teachers in monitoring the evolution/progress of students’ interactions, intervene whenever needed and obtain diagnostic information during the implementation of the appropriate assignments. The quantitative summarization of
participants’ interaction using time bubble charts for posts and comments and Social Network Analysis (SNA) for comments shows that students have been reading and commenting each other intensively during the activity. This graphical analysis of the students’ interactions during blogging is a third research dimension in Computer Supported Collaborative Learning using blogs. The work of Fessakis et al. (2008) aimed at understanding the relation of collaborative electronic learning environment (blog) and the genre of assigned work (learning by design).

Learning by design is related to constructionism, according to which new knowledge is more effectively developed by students when they are actively engaged in the construction of an external, shareable artefact that helps them to reflect and collaborate. Learning by design emphasises the learning value of the artefact design and at the same time underlines the learning benefits of the process regardless of the final product. In learning by design activities the designed artefacts are of personal significance for the students and “represent” the learning outcome. There are several views on what constitutes learning by design; according to Han and Bhattacharya (2001) learning by design environments include: authenticity of the design theme, a balanced mixture of constrained, guided, scaffolded challenges and open design tasks, a rich variety of feedback information for designers, discussion and collaboration, experimentation, inquiry and reflection.

During learning by design activities students follow repetitively, in general, the following stages: theme-question selection; target group description; design and first implementation of the artefact; pilot application of the artefact; feedback information gathering; reflection; and design adaptation. Learning by design activities foster interaction between participants (students and teachers). The teacher’s role includes supporting the students’ efforts and guiding their interaction in order to obtain the desirable learning result. The teacher needs to supervise systematically the group interactions and decide for the quality of their verbal exchanges. In a conventional classroom the students’ interaction is often not allowed at all (frontal teaching); furthermore, in the case of non-online group-teaching, teachers cannot easily monitor and intervene during the collaboration, focusing mainly on the output-product of the activity. The use of an online environment seems to be an efficient solution for managing the increased number of learning interactions occurring in group collaboration and for the improvement of the provided education quality. As was pointed out in Fessakis et al. (2008) case study, weblogs can fulfill the information management requirements of learning by design through group blogging.

Following the aforementioned work, this paper discusses/presents the quantitative interaction analysis of groups collaborating on learning by design activities through blogging.

**Interaction analysis and blogs**

Computer based Interaction Analysis (IA) is an emerging field (Dimitracopoulou, 2008) aiming at supporting directly or indirectly the participants in technology mediated activities. There are two main complementary directions of IA according to the end user of the information produced. In the first direction, the IA informs the adaptivity components of the learning environment. In the second direction, the participants have direct access to IA information supporting them to self-regulate. For example, in Fessakis et al. (2004) the authors define and test Collaboration Activity Function (CAF), a graphical collaborative interaction analysis tool that uses as raw data the actions, on a shared workspace, of the members of a small group (2-4 persons) that are collocated and solve modeling problems. The empirical examination of CAF showed that teachers can utilize it in order to a) assess collaboration; b) evaluate students’ contribution; c) choose the appropriate time for intervention; and d) assess teacher’s intervention consequences. The teachers were allowed to use CAF in order to reflect on their intervention strategies and self-regulate. Other typical works concerning interaction analysis in asynchronous communication environments include Bratitsis & Dimitracopoulou (2008) for forums and Anjewierden et al (2006, 2009) for blogs.

This paper advances the work of Fessakis et al. (2008) using a significantly larger number of groups learning by design through group blogging while focusing on the impact of graphical IA tools on participants during their collaboration. Therefore, this paper examines whether the graphical synopses of the interaction raw data of the group blogging helps the group to increase the level of awareness and to regulate its function allowing students to collaborate effectively and teachers to moderate them in a more efficient way. In the following sections the research conditions are explained along with the research questions and data collection tools, then the data is analyzed and interpreted and finally an overview discussion is provided.
Research

In the field of collaborative learning the interaction analysis is considered of great importance. IA aims at fostering productive interactions among the users (students, teachers, researchers) by increasing their awareness and facilitating their self-regulation. Besides, blogs are social software for asynchronous communication, content management systems, and recently quite popular collaborative learning environments. It is therefore useful to study IA for group blogs through which students implement learning activities according to collaboration scripts. The aim of the present research is firstly to explore the impact of IA graphs on the members of group blogs during collaboration and secondly to examine the generalization possibility of the pilot study results (Fessakis et al, 2008) concerning the consistency of blogs to the “learning by design” approach, using a larger sample. The paper focuses on blogs because they are more adequate to the “learning by design” iterative and adaptive process than other services like for instance forums or wikis. Since blogs are usually available as a component in more complex environments like Learning Management Systems, the results are interesting for learning design in general. The present research stems from the comments of participants of the previous research on graphical synopses of their action after the completion of the activity. Students showed interest in the interpretation of the graphs and commented that it would be useful if the interaction analysis graphs were available during the collaboration through the blog. Finally, the weblog technology that is available, normally, does not offer graphical representation of the interaction analysis of the users. This offers an incentive to propose and examine experimentally, specific interaction analysis graphs in order to produce design specification for future development of integrated interaction analysis toolkits for educational blogging.

Research questions

The following questions have been set for the approximation of the research aim:

**Q1.** How user-friendly are the group blogs for students and to what extent do they believe that blogs can enhance collaboration and communication skills? How consistent is blogging to the “learning design” approach? (Could the blogs efficiently facilitate the communicational and information management requirements?)

**Q2.** How useful and understandable is the information provided by the interaction analysis graphs according to the students? What graphs do the students prefer?

**Q3.** To what extent do the interaction analysis graphs affect self-regulation of the group blog members? Are there any significant differences for impact among the different kinds of graphs?

Research conditions

The experiment took place in the interval between 19 FEB and 04 JUN of 2009. The participants were 147 students from the Department of Preschool Education and Educational Design of the University of the Aegean in Greece, who were attending a course about the design and development of ICT applications in education. The majority of the students were at their third year of study while some of them were at the second or forth. The age of the students is not considered as a significant factor in this research since all of them were adults with similar educational background. Students were organized in 21 groups, consisting of 7 members each. The group allocation was made according to the list of their surnames in alphabetical order. A blog was created in a popular free blogging service for each group. The access to the blog content was restricted to the members of each group. Students participated in a three-hour training so as to familiarize themselves with blogs prior to the conduction of the research. In addition, every student on each blog had access to a manual with detailed instructions for posting, commenting, files uploading, etc. The students' task was to design technology enhanced lesson plans. The learning activity was organized according to the following phases:

**Phase 1. Socialization:** Each student publishes a brief presentation of him/her and a short reflection on the lesson plan models.

**Phase 2. First design and peer review:** Each student publishes 1-2 lesson plans on the blog and makes comments on the versions provided by the other members of the group.
Phase 3. Revision and peer review: Each student revises the lesson plans as many times he/she thinks necessary, taking into account the comments of his/her colleagues. Students continue also, the commenting of the others’ designs and revisions.

Phase 4. Final project deliverable: Finally, students compile the deliverable of the project by selecting, after discussion, the best lesson plan.

The groups were separated in three categories. The first category (K1-intragroup analysis) included the groups: (1, 2, 3, 7, 8 and 17). The researcher supplied the groups with the IA graphs (A1-A4, figures 1-4) which were summarizing data for comparison of the group members according to their contribution on posting and commenting. In addition, the groups of the category K1 had available social network analyses diagram (A5, figure 5) regarding who was commenting on whom. The second category (K2 - inter-group analysis) included the groups: (10, 11 and 19). These groups were given by the researcher IA graphs that were comparing the groups of this category to other groups in terms of total number of posts and comments (B1-B3, figures 6-8). The difference in the case of category K2 is that the members of each group were allowed to compare quantitative data of their group to other groups without reference to the individual contributions of its members. The third category (K3 – Control groups) included the remaining groups that did not receive any IA graph. The organization of the groups during the experiment and their correspondence to the different kinds of graphs is depicted in schema 1. The posting of the graphs to the blogs had been scheduled to take place after the completion of each main phase of the scenario. The graphs are presented per group category in detail in the next section.

The interaction analysis graphs of the research

The decision regarding the number and the kind of graphs was reached taking into account that most students were not experienced social software users neither were experienced graph readers. So the selected graphs provided were quite simple apart from the social network analysis diagram. Even though the SNA diagram was rather complex, it was included since students were using it in compulsory courses during their first year of study in the University.

The selection of the specific graphical representation of interaction analysis stems from the basic, visible actions of the students in the blog which are: post of an article and publishing of a comment. Weblogs, normally, do not offer graphical analyses of their users’ post and comments production. In the case of a collaboration group blog, it is useful for the members to be aware of how many posts and comments are produced by each member in order to decode simple self-regulation rules. Each group member can use this information to compare her/his contribution/productivity to the blog in relation to that of the others. The representation of this regulative information should be readable at a glance in a simple graphical format; the bar graphs of the total number of posts and comments of each member seem suitable for this purpose. Furthermore, in order to make an easy relative comparison of the number of posts and comments published by each member of the group simultaneously, we selected a bubble chart sorting the users according to the number of comments in the yy’ axis (height of the bubble center) and using diameter for the bubble proportional to the number of posts. This chart gives regulative information to the users in
relation to their reading and writing behavior in the blog. In other words each user can assess graphically whether
she/he is reading or writing more than the others through the use of a single graphical representation. Finally, there is
another important dimension of the interaction concerning not only the intention of commenting but also the variety
of the recipients’ comments. It is useful for each collaborator to read and comment to most of her/his peers. From
another perspective, it is important for the students to increase their influence on their collaborators causing
commenting to and from the most members of their group. To facilitate this kind of interaction analysis, we selected
SNA diagrams constructed with a vertex for each user and a directed arc for each pair that has commenting users, the
arrow of the arc depicts the receiver of the comments. The thickness of the arcs is proportional to the total number of
comments while the exact number of comments is mentioned explicitly as a label to the arc near the recipient vertex.
This complex diagram gives information to the users in order to become aware of how many other users (and which)
they read and help with their comments. Furthermore students can see how many of their peers are interested in their
work and who worked harder than the others in the group.

In fact, the interpretation of the diagrams depends on several factors, e.g., the phase of the learning scenario, the
content of the posts and comments, the kind of comments (positive, negative). In any case simple self-regulating
rules are expected to be derived from the availability of graphs. More specifically, the post bar graphs during the
lesson plans proposal is expected to encourage/motivate students with the least number of proposals to publish some
more while the number of comments posted bar graph during the peer review phase could guide the students to read
and review more lesson plans from their colleagues. The ‘comments received’ bar graph in combination to their
content gives an indication to the students regarding the quality of their proposed lesson plans and their influence (on
what?). Finally the SNA diagram could lead students to increase their circle of collaborators reading and
commenting those who have not yet (done what?).

As far as the inter-group interaction analysis is concerned similar graphs are selected for the posts and comments
summarizing and comparing whole group production instead of separate users. In this instance SNA is not
applicable.
In any case, the IA graphs should depict information for arbitrary time intervals of the learning process (e.g., each
phase of the learning scenario) in order to inform students and teachers about the dynamics of the collaborators’
contribution. The time periods are related loosely to the phases of learning scenario because of the semiautomatic
way of their production by the researches. The production of graphs for each phase facilitates its interpretation
according to the semantics of the phase, e.g., much commenting in the socialization phase has different meaning than
in the peer review phase. In a fully developed and integrated interaction analysis system, students and teachers
should have the opportunity to choose arbitrary time intervals and produce automatically any interaction analysis
graph they choose at any time. Based on the aforementioned assumptions specific diagrams are assigned to group
categories to support intragroup and intergroup interaction analysis as it is described in detail in the next sections.

*Graphs for the groups of category K1*

The interaction analysis graphs enabled students to retrieve information about their own actions as well as those of
the members of their group. More specifically the graphs for category K1 groups were: **Graph A1:** Bar chart of the
number of posts of each member per period and in total (Figure 1). **Graph A2:** Bar chart of the number of comments
published by each member per period and in total (Figure 2). **Graph A3:** Bar chart of the number of comments
received by each member per period and in total (Figure 3). **Graph A4:** Bubble chart of the number of posts and
comments published by each member in total (Figure 4). The height of the bubble is proportional to the number of
comments while the diameter is proportional to the number of posts. **Graph A5:** Social Network Analysis diagram
(Figure 5). The nodes represent the students while the arcs depict commenting.

*Graphs for the groups of category K2*

Regarding the groups of the K2 category (10, 11, 19) the graphs were comparing the number of posts and comments
of each such group to those of category K1 groups (1, 2, 3, 7, 8, 17). More specifically the graphs in this case were:
**Graph B1:** Bar chart of the number of posts of each group per period and totally (Figure 6). **Graph B2:** Bar chart of
the number of comments of each group per period and totally (Figure 7). **Graph B3:** Bubble chart of the number of
posts-comment for each group (Figure 8). The height of the bubble is proportional to the number of comments while the diameter is proportional to the number of posts.
Research data collection

As far as questions 1 and 2 were concerned, the data was collected using online questionnaires. With reference to question 3 the raw data about the students’ actions to the blogs was analyzed. This data was used for the production of the interaction analysis graphs by a semi-automatic process. The process of the specific graphs production could be fully automated if there was full access to the blog platform database. In this study we chose to use a free popular weblog service because it is a study of feasibility, user requirements specification and verification of the proposed graphs. The full development of the automatic production of the graphs and their integration in the blogging service could be a future elaboration of the work.

In addition to the quantitative data of user actions logging, the students of the category K1 and K2 groups were asked to comment on the posts with the graphs on their blogs in order to (a) make sure that they had noticed the graphs and (b) collect evidence about what information they decode from them.

Research data analysis – Results

Q1. How user-friendly are the group blogs for students and to what extent do they believe that these blogs can enhance collaboration and communication skills. How consistent is blogging to the “learning design” approach? (Could the blogs efficiently facilitate the communicational and information management requirements?)

Based on the students’ answers to the questionnaires at the end of the experiment we can conclude that: (a) the students had familiarized themselves with the blogs within a few hours. Publishing articles and comments, blog browsing and monitoring, was easy for the students while many of them found the file uploading quite complicated in the beginning for the specific blogging service, (b) 58% of the students felt that group blogging helped them to improve their collaboration and communication skills, (c) 70% of the students stated that they would use blogs again in a future collaboration. Considering the above, it can be concluded that blogs are not only easy for students to use but they also fulfill the requirements of online group “learning by design” scenario to a satisfactory extent. As was rather expected, there is no statistically significant differentiation among the group categories. The results are consistent to those of the pilot study (Fesakis et al., 2008), supporting their generalization in similar situations.

Q2. How useful and understandable is the information provided by the interaction analysis graphs according to the students? What graphs do the students prefer?

Category K1 groups’ students’ views

With reference to the question, “Which graph did you take most into account?” 52.38% of the students of category K1 groups answered that they took into account the graph A4 (posts-comments bubble chart), the second most popular graph being A1 with percentage of 33.33% and the least popular being the SNA diagram A5 (9.52%) and the comments bar graph A3. As to the question, “How useful is the information of the graphs?” the students evaluated all the graphs characterizing the information from “enough” to “very useful” according to the following percentages: A1 – (>85%), A2 – (>85%), A3 – (~70%), A4 – (>85%), A5 – (~70%). In the case of SNA (A5), 9.52% of students answered that the information was not useful at all. In the question, “How much did the graphs affect your behaviour-participation to the blog?” more than 70% of the students answered from “fairly” to “very much” for each of the graphs A1, A2 and A4, while the percentage for A3 is ~60% and for the SNA diagram A5 ~65%. In the case of SNA, 4.76% of students answered that they were not affected at all by this graph. This is significant because no student selected this answer for any of the other graphs. At the same time, SNA received the largest percentage (14.29%) of students answering that they were affected considerably. No other graph gathered so many answers in the “very much” option. This could be due to the fact that most of the students were not able to interpret the rich information of the SNA while those who were able to do so were affected by SNA the most. The students that were not able to interpret the SNA diagram could not use it to make a self-regulation plan. This is easy for the bar charts where every student can understand rather straightforward if she/he needs to increase his contribution. The comments of the students on the SNA diagram post that follow are informative and show that some students could use SNA as brief representation of the “dynamics” and the “liveliness” of their group:
Student K. O. said: The SNA diagram show the dynamics of our group. I think that our group function is satisfactory!!! 02 April 2009 10:26

Student K. E. said: The SNA diagram reflects the dynamics of our group. A group which works in a good tempo. 08 April 2009 12:05

Student K. E. said: The SNA express the liveliness of the participants to the group.....08 April 2009 10:13

Similar appear to be the answers of students to the question, “How clear-understandable is the information of the graphs?” For the graphs A1-A4 more than 90% of students selected the “fairly” or “very useful” answers. In the case of SNA diagram the same percentage decreases to ~70%, while 23.81% answered that SNA is “little” understandable and 4.76% answered “not at all.” This explains more the findings in the research question 1 and justifies the selection of simple graphs in this case. Users need training on information decoding and guidelines for the transformation of this information to regulative action in order to use more complex interaction analysis graphs. In the complementary question, “What graphs need guidelines for their reading?” students answered: A5 – SNA diagram (42.86%), A4 – bubble chart (33.33%) and then A1-A3 (< 5%). A percentage of 28.57% of students answered that there is no need for guidelines for reading the graphs. Summarizing, students seems to take most into account, for self-regulation, the graphs concerning the number of posts per student and the number of comments he/she had received (A4).

Category K2 groups’ students’ views

With reference to the question, “Which graph did you take most into account?” students answered: B1 – (36.36%), B2 – (9.09%) and B3 – (54.55%). Consequently, the bubble chart of posts and comments (B3) had the greater impact. In the complementary question, “Which graph did you take least into account?” students answered: B1 – (36.36%), B2 – (63.64%) and B3 – (0%). The answers are consistent to the previous outcome supporting the reliability of the data. In the question, “How useful is the information of the graphs?” students answered “fairly” or “very much” in a percentage > 90% for each of the graphs B1-B3. In the question, “How much did the graphs affect your behaviour-participation to the blog?” the students answered “fairly” or “very much” in ~55% for B1 and ~70% for each of B2 and B3. This is reasonable since B1 depicts the number of posts for which the least number was defined in the scenario and the groups did not have significant variations. In the question, “How clear-understandable is the information of the graphs?” the students answered “fairly” or “very much” for all the graphs. In the complementary question, “What graphs need guidelines for their reading?” the students answered: B1 – (9.09%), B2 – (18.18%), B3 – (45.45%) and «none» – (36.36%). The percentages are proportional to the complexity of the graphs.

Q3. To what extent do the interaction analysis graphs affect self-regulation of the group blog members? Are there any significant differences for impact among the different kinds of graphs?

In question 3, we examine the impact of IA graphs to the groups of students collaborating through the blogs. For the comparison of the group categories we use: (a) the mean number of posts and (b) the mean number of comments published, per group for each category. The data collected from the blogs show that the total number of actions for the 147 students was greater than 2095 (817 posts and 1278 comments). The comments of students to the graphs posts are not included in the calculation. Tables 1 and 2 summarize a separate variable for comments and posts for each group category. There are six variables of the form: KXC for comments and KXP for posts where X denotes the category of group (X in \{1, 2, 3\}). K1 and K2 categories are similar except from the kind of graphs they had available. Consequently any significant differences to the control groups’ category K3 and among K1 and K2 could be related to the impact of the interaction analysis graphs.

In order to examine whether there is any significant difference among the groups of the different categories we could apply ANOVA using the category (K) as a factor variable and the Comments (C) and Posts (P) variables as dependent variable. In order for the ANOVA to be applicable, there are specific prerequisites that need to be met: The samples should be random and independent and the variances of the populations should be equal. In our case the samples are fairly random because of the use of the surname list on alphabetical order for the group formation. In addition the samples are independent because they have not got students in common.
Table 1. Basic statistics for the comments variables per groups

<table>
<thead>
<tr>
<th>Var</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>μ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1C</td>
<td>6</td>
<td>72</td>
<td>173</td>
<td>111.83</td>
<td>38.52</td>
</tr>
<tr>
<td>K2C</td>
<td>3</td>
<td>56</td>
<td>74</td>
<td>66.33</td>
<td>9.29</td>
</tr>
<tr>
<td>K3C</td>
<td>12</td>
<td>10</td>
<td>69</td>
<td>34.00</td>
<td>18.85</td>
</tr>
</tbody>
</table>

Table 2. Basic statistics for the posts variables per groups

<table>
<thead>
<tr>
<th>Var</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>μ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1P</td>
<td>6</td>
<td>40</td>
<td>57</td>
<td>49.33</td>
<td>6.41</td>
</tr>
<tr>
<td>K2P</td>
<td>3</td>
<td>39</td>
<td>53</td>
<td>45.33</td>
<td>7.09</td>
</tr>
<tr>
<td>K3P</td>
<td>12</td>
<td>0</td>
<td>47</td>
<td>32.08</td>
<td>8.68</td>
</tr>
</tbody>
</table>

Furthermore, the distribution of the general populations are probably normal as the Shapiro-Wilk (S-W) normality suggests: \( S-W_{K1C} = 0.927; p = 0.554, a = 0.05 \), \( S-W_{K3C} = 0.937; p = 0.461, a = 0.05 \), \( S-W_{K1P} = 0.953; p = 0.762, a = 0.05 \) and \( S-W_{K3P} = 0.963; p = 0.831, a = 0.05 \). For the K2C and K2P normality it is not possible to apply a formal test because of the small number of groups (N < 4) but we can hypothesise this by taking a reasonable risk or overriding it using non-parametric tests. The last prerequisite of ANOVA for the variance equality is checked using Levene’s test. More specifically for the comments variables we apply, Levene’s test (Median) \( (K1C-K3C) = 4.141, df1 = 2, df2 = 18, p = 0.033, a = 0.05 \) therefeer the hypothesis \( H0: The \ variance \ are \ identical \) is rejected and the prerequisite is not fulfilled. In contrast, for the case of posts variables using the same test we apply Levene's test (Median) \( (K1P-K3P) = 0.411, df1 = 2, df2 = 18, p = 0.669, a = 0.05 \). Consequently we have to accept hypothesis \( H0: The \ variances \ are \ identical. \) This is quite reasonable since the minimum number of posts was defined by the learning scenario and the students were not significantly differentiated. The above mean that for the case of comments variables we should use a non-parametric test like Kruskal-Wallis while for the post variables it is possible to use ANOVA. So we have Kruskal-Wallis \( (K1C-K3C) = 14.628, df = 2, p = 0.001, a = 0.05 \) which means that we should reject \( H0: The \ samples \ come \ from \ the \ same \ population, \) and Kruskal-Wallis \( (K1P-K3P) = 11.433, df = 2, p = 0.003, a = 0.05 \) as well as for ANOVA \( (K1P-K3P) \) where \( F = 10.586, sign. = 0.001 < 0.05 \). All the tests show significant differences among comments and posts variables mean according to the group category factor.

At this point we will examine which pairs of variables show significant variation. We reached the conclusion that the groups of K1 category produced on average more posts and comments than the control groups of category K3 as well as from the groups of category K2 (significant differences on mean according to Mann-Whitney test for \( a = 0.05 \)). Furthermore, the groups of category K2 produced also more comments and posts on average than the control groups. Especially for the mean number of posts we have significant differences according to t-test only among K1P-K3P and K2P-K3P. The evidence of the research support the hypotheses that interaction analysis graphs had a positive impact on the intensity of group collaboration in terms of the number of posts and comments they produced. In addition, the graphs for the category K1 (intragroup analysis) seemed to have a stronger impact in terms of mean number of comments than those of category K2 (inter-group analysis). This could be the case because intragroup interaction analysis graphs compare the collaborators’ contributions and thus increase the competition among the students. The following excerpt from the comments of a member from group 8 for the graph A4 is revealing:

K. E. said: This graph constituted by circles showing how much a student deal with the group. And it seems that I am the least working member!!! I will agree with the other members of the group that this graph as well as the previous provides an incentive for me to try harder. 08 Apr 2009 10:19

Taking into account all the above it could be argued that a more efficient way in helping students engage in collaboration and self-regulation, would be to provide them with graphs of both categories.

Conclusions and future work

The continuously increasing use of blogs in education and especially in the context of Computer Supported Collaborative Learning scenarios makes their systematic educational study extremely interesting. Blogs are quite easy for students to use while they fulfill at a satisfactory level the communication and information management.
requirements of online collaborative learning by design scenario. The interaction analysis graphs of collaborating
students could help teachers to monitor, moderate, coordinate, assess, etc., and students to increase their awareness
and self-regulate during their participation. The interaction analysis is also interesting for education researchers. The
presented study advances on a previous pilot study using a larger sample of participants to examine their views on (a)
the usability of blogs, (b) the suitability of blogs for the implementation of the “online collaborative learning by
design” scenario, and (c) the facilitation of general communication and collaboration skills by blogging. The research
results are fairly positive and consistent to those of the pilot study. More specifically the research data confirms that
most students learn to use blogs easily, they recognize the suitability of blogs to the communication and information
management requirements of the scenario and believe that blogging in the particular context of the learning scenario
makes them better collaborators.

Furthermore the research explores the impact of interaction analysis graphs organized in two categories. The first
category includes: (a) bar graphs and a bubble chart that summarize the evolution of the contribution in posts and
comments for each group member separately, and (b) a Social Network Analysis diagram with one node per student
and arcs showing who was commenting on whom. For each group the SNA depicts how extended and intensive the
communication (in term of comments) developed was during collaboration, permitting the analysis of this
communication for each pair of students. The graphs of the first category concern intra-group interaction analysis
and aim to facilitate competition among the group members. The second category of graphs includes bar graphs and
bubble chart summarizing the volume of posts and comments of whole groups, thus permitting the comparison
among groups. The graphs of the second category concern intergroup interaction analysis and aim to facilitate
competition among the groups. From the experiment presented in the paper there is evidence of statistically
significant differences among groups, posts and comments production depending to the presence and the category of
graphs. This means that the interaction analysis graphs that were used had a significant impact on collaborative
groups helping them to self-regulate during the learning scenario implementation in order to increase their
participation and intension of their efforts. As far as the different categories of graphs is concerned (intra vs. inter
group analysis), in this study the intra-group analysis graphs had a significantly stronger impact resulting in more
active, productive and engaged groups with extensive commenting among their members. The Interaction Analysis
Graphs give students the impression that teachers are monitoring their participation in the groups and this
facilitates/urges students to contribute and collaborate more.

The current research results which were based on a large scale experiment support the generalization of the claim of
the previous case study (Fessakis et al., 2008) that collaborative learning by design model is supported efficiently by
weblogs in terms of information management and communication requirements for teachers and learners. In addition
the research validates the efficiency of the proposed interaction analysis graphs on enhancing learners’ self-
regulation and makes their integration in educational blogging systems a justified goal.

The main weakness of this research can be identified as the small sample of the groups of category K2 which does
not allow a secure conclusion for the results concerning the graphs of the specific category. In contrast, the results
from the comparison among the interaction analysis graphs of category K1 to the control groups are quite strong.
Future aims of the research include the automatic production of the graphs and their smooth integration in the
learning environment in order to give continuous access to the users and be possible to collect interaction data of the
students to the graphs. More specifically the research supports the feasibility of an Interaction Analysis toolkit in the
form of a plugin that could be integrated in learning environments that educational blogs offer. Teachers and students
could utilize this toolkit in order to support self-regulation and monitoring of the collaboration process to improve
learning in design scenarios. The toolkit will permit the dynamic production of interaction analysis
graphs for an arbitrary user at selected time intervals upon demand and could be accompanied to an intelligent
system to support their interpretation in self-regulating rules.

References

http://www.lifewithalacrity.com/2004/10/tracing_the_evo.html

eexample. Lecture Notes in Computer Science, 4277, 279-289.


Can Speaking Activities of Residents in a Virtual World Make Difference to Their Self-Expression?

Soojeong Lee
Kyungnam University, 449 Wolyeong-dong, Masan City, South Korea // 2leesoojeong@gmail.com

(Submitted June 20, 2011; Revised November 13, 2011; Accepted January 09, 2011)

ABSTRACT

The purpose of this study is to search for any difference in self-expression of Second Life residents with different levels of shyness. For this purpose, we used sixty students from two fifth-grade elementary school classes. Thirty students were assigned to the high shyness group and the rest to the low shyness group. Each group completed pre- and post- self-expression tests. After six weeks of speaking activities in Second Life, the results indicate that self-expression scores increased for students in both the high and the low shyness groups. The low shyness group showed an increase by 1.00 in the self-expression score. However, the high shyness group showed an increase by 3.14 after the speaking activities. This result suggests that Second Life can be a good environment to enhance self-expression in students, especially those with degrees of high shyness.

Keywords
Shyness, Second Life, Virtual world, Self-expression, Speaking activities

Introduction

Shyness is the feeling of apprehension, lack of comfort, or awkwardness (Crozier, 2000; Carducci & Golant, 1999) experienced when a person is in proximity to, approaching, or being approached by other people, especially in new situations or with unfamiliar people. So, shyness can hinder people’s performance in many aspects of life. It can make people reluctant, make them afraid, or even create negative feelings toward others. In return, this causes those involved to feel even more shy and to blame themselves. In these situations, if others insult or coerce them, they may feel anger or become critical of others (Bortnik, Henderson, & Zimbardo, 2002). In classrooms, student shyness can result in students saying nothing and not expressing themselves. In addition, shyness is found to have a negative correlation with school achievement, undermining school achievement (Jeong, 1997). What’s more, shyness triggers various problems such as depression, low self-esteem, low social and emotional adjustment, and low self-efficiency (Shin, 2002; Yoo & Park, 2007; Lee, 2005). Shyness has become correlative with having a low self-esteem (Crozier, 1995), as most withdrawn people see interaction and active lifestyle as pointless. It should not be surprising that shyness is associated with a host of negative adjustment outcomes, including internalizing problems (e.g., low self-esteem, loneliness and anxiety) and poor school adjustment (Arbeau, Coplan and Weeks, 2010).

In particular, what stands out among shy students is that they are unwilling to express themselves. This is evident in a number of studies focusing on the effect of self-expression classes targeting shy students (Goh, 2005; Baek, 2003; Ha, 2006). Self-expression represents an ability to convey one’s own emotion or to express one’s own intention (Kim & Lee, 2003). A study (Kim, 2004) shows that self-expression contributes to upgrading sociability. Given this, it may be observed that self-expression serves as the basis for socialization. Just as education ultimately aims to improve learners’ sociability, self-expression is an essential personality trait achieved through school education. Despite that, schools fail to pay attention to promoting children’s self-expression in class activities due to too much emphasis on the cognitive aspect of the lesson. This makes shy students afraid to speak up, to be distressed due to low self-esteem and poor sociability, allowing their negative emotions and way of thinking to have an ill effect on others. These observations suggest that shy students require educational concern, assistance, and consideration.

The studies by Goh (2005), Baek (2003), Yee (1992) and Ha (2006) focus on developing ways to promote self-expression in classroom interactions. These studies use an assertive training as a treatment for decreasing shyness of students. Moreover, these studies identify the effect of assertive training on shyness in classroom situations. The treatment of these studies is a specially designed program for their experiment. In their nature, they are not activities which students perform in their daily lives. In addition, these studies focus on controlling learner variables to promote self-expression by treating shyness and are limited in that they fail to provide any effective ideas about controlling the classroom variables that might contribute to shyness. If classroom variables such as
instructional procedures, media, organization, and environment, still remain out of control, they may trigger shyness among students found to report reduced shyness as a result of taking self-expression programs.

This raises the need to develop ways of promoting self-expression in classroom interactions by reducing the ill effects of classroom variables on shyness. That is where cyberspace and anonymity can come to the fore. Kim (1999) argues that the psychological sense of liberation in cyberspace helps people to loosen the shackles of their emotions and self-expression, allowing them to express themselves more strongly and exaggeratedly. Kim (2004) contends that as a result of conducting in-depth interview on adolescents, 14 out of 26 respondents were found to act differently in cyberspace. A large portion of the respondents were found to be more talkative and less afraid of speaking up in cyberspace. In Second Life, there are four main forms of communication: chat, instant messaging, gestures and voice. A primary benefit of text-based communication in Second Life is that students and tutors can have a transcript of what they have discussed during the session, so they revisit later on. It is also useful for those students who were unable to participate in the Second Life session to have an idea of what happened. Voice-based offers advantages over text-based if the nature of the communication is for providing feedback or talking things through, etc.

In addition, Reid (1991) says that guaranteed anonymity in cyberspace gives people a chance to disguise or conceal what they are, promoting intense socialization hardly ever seen in real interpersonal relationships. Anonymity is a concept that evokes an absolute lack of connection between a piece of information or an action and a person (Wallace, 1999). In other words, one has anonymity or is anonymous when others are unable to relate a given feature of the person to other characteristics. Thus, anonymity presupposes social relations and relative to social contexts. For this reason, anonymity affects social behavior in cyberspace as well as in real life. Park and Kim (2000), whose studies focus on the effect of anonymity on discussion in synchronous virtual discussion class, say that anonymity contributes to more attempts at self-expression. Jeong (2004) says that the anonymity guaranteed in online discussion results in an increase of messages exchanged in terms of sociability and interaction. These studies strongly support that residents in Second Life can have their own behavior affecting another lives.

All of these studies show that cyberspace, and the anonymity it provides, may contribute to more attempts at self-expression. This study aims to identify the accurate effects of shyness on other students’ self-expression and the effectiveness of the classroom variables provided. By verifying the effectiveness of classroom variables, this study can inform teachers of applicable classroom variables that promote self-expression during class, particularly for shy students. This will ultimately help to boost learners’ school achievement and sociability.

This study focuses on identifying any difference in self-expression by introducing cyberspace self-expression boosters to the elementary school environment. In doing so, this study aims to shed light on the effectiveness of cyberspace self-expression boosters as a classroom variable that contributes to controlling shyness in classroom interaction. To that end, this study focuses on the research question described below.

Is there any difference in self-expression of residents with different levels of shyness after they perform speaking activities in Second Life?

By identifying the effectiveness of classroom variables, this study can inform teachers of applicable classroom variables that promote self-expression during class, particularly for shy students. This will ultimately help to boost learners’ school achievement and sociability.

Limitations of the study

This study does not aim to clarify the casual relationship between speaking activities and self-expression of residents in Second Life. Instead, this study focuses on different outcomes of self-expression originated from student’s shyness in Second Life. In order to see any causal relationship, an experimental research design should be used. However, it seems irrelevant to set up a control group where treatment effects are expected with the students who participated voluntarily in the study. Performed within this context, this study has a limited generalization of results.
Second Life as a doing place

Second Life (http://www.secondlife.com) is a virtual world developed by Linden Lab and is accessible via the Internet. The Second Life Viewer enables its residents to interact with each other through their avatars. Second Life users create a customized avatar or personage to represent themselves (http://secondlife.com/whatis/). The avatar moves about in the virtual world using mouse control and intuitive keyboard buttons. They can meet other residents, socialize, participate in individual and group activities. In addition, they create and trade virtual property and enjoy pastimes, or travel throughout in-world. Second Life’s virtual world also includes sound; wind in the swaying trees, babbling brooks, audible conversation, and built-in chat and instant messaging. Residents buy property, start businesses, game with other residents, create objects, join clubs, attend classes, or just hang out. Residents can also travel freely throughout the virtual world. Thus Second Life is a place for residents’ communication and expression by doing various activities.

Using Second Life in education can be extremely advantageous in several aspects (Bransford & Gawel, 2006). First, it creates a sense of sharing amongst residents meeting online. It supports to create a strong sense of collaborative community. Second, Second Life provides an environment that expands resident’s ideas through interoperability. In Second Life, participants can merge objects with other things built within Second Life. Third, Second Life makes it possible to create interactive learning experiences that would be hard to duplicate in real life. Based on these characteristics, there have been efforts of to use Second Life in teaching and learning. Bilyeu (2007) reported that the students’ responses to using Second Life were very reassuring in that they found playing these experiences a fun way to learn the science material.

On the other hand, Second Life has certain implications for ‘learning by doing.’ Throughout history, youths have been apprenticed to masters in order to learn a trade. We understand that learning a skill means eventually trying our hand at the skill. When there is no real harm in simply trying, we allow novices to "give it a shot" (Schank & Clearly, 2011). Traditional educational approaches which focus on knowledge acquisition have been criticized for disregarding self-initiated learning activities for students. The idea of ‘learning by doing’ emphasizes meaningful learning and the integration of various aspects of real life. In other words, ‘learning by doing’ links subject matter to substantial experience for students through their active participation and/or work. “Learning by doing” in Second Life has the following implication for teaching and learning:

- By observing how things occur in the world and trying activities on their own, student learn to trust their own abilities.
- By creating and re-inventing, authentic experiences can grow.
- By interacting with others, residents can increase collaboration.
- By volunteering, residents can dedicate their time to becoming part of the learning community.

‘Learning by doing’ is similar to the way parents usually teach children. They don't give their children series of lectures in order to teach them how to walk and talk. When it comes to school work, however, instead of encouraging students to learn by doing, we create courses of instruction that merely deliver information, knowledge and concepts to students. In Second Life, learning by doing is natural for the residents.

Virtual worlds and self-expression

Self-esteem plays an important role in promoting self-expression (Kim & Lee, 2003). Self-esteem may be described as self-efficacy. Bandura (1986) says that self-efficacy may increase with the use of resources boosting efficacy expectancy such as accomplishment experience, vicarious experiences, verbal persuasion, and emotional arousal. Among them, accomplishment experience is the most influential determinant which decides the assessment of self-efficacy. The higher one’s accomplishment experience, the higher their self-efficacy. But when one experiences failure repeatedly, the opposite is true. Elements such as competence on performance accomplishment, performance complexity, outsider’s help, situation at the time of performance, and effort all play a part in assessing self-efficacy.

Bandura (1986) proposed effective models to induce successful performance by participants. In those studies, participants were well informed of what should be done. The models include performance desensitization that allows a gradual approach to performance accomplishment. Participant modeling allows both observers and
demonstrators to play the given roles successfully while granting exposure to accomplished performance situation and self-fulfillment of performance accomplishment.

The anonymity and indirect self-expression in virtual worlds may allow shy students, afraid of expressing themselves, to get more used to speaking up in front of learners, more so than in face to face meetings. This may be called a sort of performance desensitization proposed by Bandura (1986).

Cyberspace anonymity contributes to desensitizing shy students in accomplishing performance. If cyberspace anonymity is introduced as a classroom variable, it is expected to boost learners’ sociability and to promote their self-expression. Kim (1999) explains how cyberspace anonymity can result in promoting self-expression in classroom interactions. Firstly, Kim (1999) argues that the psychological sense of liberation in cyberspace helps people to loosen the shackles of their emotions and self-expression, allowing them to express themselves more strongly and exaggeratedly. Secondly, openness and equality play a part. In addition, a cyberspace gives its residents infinite freedom, which transcend human subjectivity and where identity becomes no longer burdened by the prejudices of persons (Miah, 2000). All users are treated equally or in a less hierarchical way in a cyberspace. In Second Life, when residents live an ordinary life, they are given the freedom to act or say whatever they like, not only that, they are also given the opportunity to be whom, or whatever they like. This allows residents to live out their fantasies, or sometimes even real lives anyway that they choose. The best part about all of this is when they are interacting in Second Life then they don’t have to worry about people judging the real you, they are simply analyzing the avatar, so they no longer have to be self-conscious about talking to new people. The only limit that residents will have is their own imagination. This sense of equality is reinforced in cyberspace, leaving users much less exposed to authoritarianism. Thirdly, complex identity comes into being. People hold their own identity in diverse ways. Cyberspace anonymity allows people to be selective about when and where they expose their identity or even to disguise it. This gives users latitude to present their identity in various ways and gives them freedom to show themselves as they want.

Anonymity and indirect exposure of identity in cyberspace contribute to the desensitization of shy people to accomplishing performance. In addition, by introducing cyberspace anonymity to the classroom, learners may get used to speaking up without difficulty. This enables learners to express themselves more often and to gain satisfactory presentation experiences. By repeating successful self-expression experiences, shy students may well experience improved self-expression and self-efficacy (Kim & Lee, 2003; Noh, 2000; Chang & Jang, 2008).

Method

Participants

Seventy-one students from two fifth-grade elementary school classes in Masan, South Korea participated in this study. They consisted of thirty-five male and thirty-six female students. Eleven students were excluded from this study due to their incomplete scales. Sixty students were used as final subjects for analysis as shown in Table 1. Half of sixty students were assigned to the high shyness group while the other thirty students were assigned to the low shyness group.

<table>
<thead>
<tr>
<th></th>
<th>High shyness group</th>
<th>Low shyness group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Tools

Two measures were administered to the participants: The Revised Cheek and Buss Shyness and Sociability Scale, and the assertiveness scale developed by Rakos & Schoeder (1979).
The Revised Cheek and Buss Shyness and Sociability Scale

The revised Cheek and Buss shyness and sociability Scale (Cheek & Buss, 1981) was used to test the shyness of the participants. This scale consisted of thirteen Likert-scale items. Participant responses are summed up as a degree of shyness, thus the higher the score, the higher the degree of shyness. The minimum score possible is 13 and the maximum score is 65. According to Cheek (1981), scores higher than thirty-three can be classified as high shyness. In this study, thirty-four was used as the border between the low shyness and the high shyness. Cronbach’s α is .90 (Heiser, Turner, & Beidel, 2003).

The self-expression scale

Participants’ self-expression was tested using Rakos & Schoeder’s (1979) Self-Administered Assertiveness scale. This scale consisted of twenty questions, of which nine questions rate assertive behavior, seven rate assertive voice, and four rate assertive physical expression. This study excludes the four physical expressions of assertiveness. The test-retest reliability of this scale is .758. The score ranges from 0 to 16.

Procedure

Participants were allocated into either the high shyness group or the low shyness group according to their shyness level measured by the revised Cheek and Buss Shyness and Sociability Scale. The mean of shyness is 40.12 and the standard deviation is 5.60. Sixty students were divided into two groups. The median was 43.24. The higher half of the sixty students was labeled as the high shyness group and the lower half was labeled as the low shyness group. The pre-test of self-expression was administered prior to speaking activities in Second Life. The mean of the pre-test of self-expression was 10.61. The post-test of self-expression was administered after speaking activities in Second Life. The mean of post-test of self-expression was 12.68.

Selection of speaking topics in Second Life

A total of sixteen topics were selected for participants’ speaking. These topics were selected from language, ethics, and social studies areas.

Speaking activities in Second Life

For speaking activities, two classes are used to form eleven groups with members of six to seven members each. The most important consideration in grouping was to maintain anonymity of the members. To maintain anonymity, unfamiliar or unknown members were put into the same group. It was forbidden to call anyone’s name; instead Second Life identities (avatar names) were the only names that could be used during the speaking activities. Two Second Life classes were held at the same time so that no subjects could be recognized. Once the speaking activities were over, the subjects were asked if they could recognize someone in the same group. All members answered that they were not able to recognize anyone. The speaking activities in Second Life were held twice a week for six weeks with each session lasting 40 minutes. In total, subjects had 480 minutes of speaking activities. Speaking topics were copied from activities from fourth-grade language, ethics, and social studies lessons. Participants wrote their opinions of a topic in their journals. They were allowed to see their memos while they are speaking to others. Even though memos are not directly related to their speaking, memos can affect their speaking time. Thus, 15 lines were permitted for their memos to help them keep the amount of speaking time the same for all members. In Second Life, members sat in their chairs while the speaker stood in front of a podium and spoke in front of other members. Because they were not allowed to talk to anybody before speaking in Second Life, they could not identify who is the speaker in real world. After all members spoke, they participated in a question and answer session for ten minutes. They discussed unclear points, speakers’ good points, and described points that need to be avoided when speaking. Figure 1 shows a subject speaking in Second Life and Figure 2 shows group discussion after speaking.

The researcher, as a moderator, moved around the classroom in Second Life, helped the subjects to manage their environment, and assisted them in using their audio equipment (e.g., microphones and headphones). All notices to
participants were delivered through chats.

**Figure 1.** Speaking in Second Life

**Figure 2.** Discussion after speaking

### Data Analysis

This study assumed an increase in the self-expression score and attempts to discover which groups show an increase on the self-expression score. To obtain basic statistics, means and standard deviations on the pre and post self-expression test for each shyness group were calculated. Then, gain scores on the self-expression test before and after the treatment were obtained by subtracting scores of the pre self-expression test from those of the post self-expression test. An independent t-test procedure was applied to test differences in increases in the self-expression of the two shyness groups, because two scores are independent.

### Results and Discussion

To examine the difference of the self-expression score after speaking activities of the residents, pre and post test scores were compared between the two shyness groups. The results are as follows.

**Table 2.** Basic Statistics

<table>
<thead>
<tr>
<th>self-expression test</th>
<th>group</th>
<th>M</th>
<th>SD</th>
<th>N</th>
<th>df</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>High shyness</td>
<td>9.13</td>
<td>2.89</td>
<td>30</td>
<td>58</td>
<td>-4.200</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Low shyness</td>
<td>12.10</td>
<td>2.58</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>High shyness</td>
<td>12.27</td>
<td>2.18</td>
<td>30</td>
<td>58</td>
<td>-1.473</td>
<td>.146</td>
</tr>
<tr>
<td></td>
<td>Low shyness</td>
<td>13.10</td>
<td>2.20</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 contains basic statistics of the pre and post test scores of each shyness group. On the pre-test, the high shyness group’s mean score was 9.13 and the low shyness group’s was 12.10. The low shyness group shows higher score on the pre self-expression test than the high shyness group. This difference is statistically significant ($t = 4.200, p < .001$). As shown in Table 2, the high shyness group has a lower level of self-expression, whereas the low shyness group has a higher level of self-expression. This finding coincides with recent studies reporting that extrovert students of less shyness tend to express themselves frequently in their web-based learning and ordinary daily life (Kim & Lee, 2003; Noh, 2000; Chang & Jang, 2008). This higher degree of expression frequency might be because less shy students are more active and express themselves more than others. On the other hand, shier students are more passive and express themselves less than others.

On the post-test, the high shyness group’s mean was 12.27, which represents an average increase of 3.14 from the pre-test. The low shyness group’s mean was 13.10, which represents an average increase of 1.00. Table 3 shows the difference in gain scores.

As shown in Tables 3, the high shyness group showed an increase in self-expression by 3.14, whereas the low shyness group showed an increase of 1.00. This difference is statistically significant ($t = 5.217, p < .001$). Thus we found the 2.14 difference in the increase of self-expression score between the high shyness group and the low...
shyness group after having performed speaking activities in Second Life. Figure 3 shows a graphical representation of the difference.

| Table 3. The increase of self-expression before and after the treatment |
|--------------------------|-----------------|--------|-----|--------|-----|
|                         | group           | M     | SD  | N     | df  | t     | p   |
| Increase of self-expression | High shyness    | 3.14  | 1.98 | 30   | 58  | 5.217 | .000 |
|                         | Low shyness     | 1.00  | 1.05 | 30   |     |       |     |

Figure 3. Increases of self-expression for high shyness group and low shyness group

This figure suggests that speaking activities in Second Life can work on shy students. Participation in the Second Life speaking activities enabled the high shyness group to achieve the same level of self-expression as the low shyness group.

Virtual worlds, including Second Life, can be used as an environment for enhancing self-expression via speaking activities. In classroom settings, shy students cannot easily make their opinions known to the teacher or to other students even though it is critical in achieving classroom objectives. As a result, shy students have difficulty in achieving learning goals in the classroom. A virtual world like Second Life with relevant activities provides them with expanded opportunity to achieve their learning goals. These activities are more effective to the highly shy students. Even for low shy students, these activities are effective even if it is not statistically significant. What this tells us is that virtual worlds can enhance residents’ self-expression regardless of their levels of shyness. Providing students with relevant activities with anonymity in a virtual world is an efficient way of overcoming their shyness.

Conclusion and suggestion

The differences of self-expression between students who are highly shy and students who are low shy exist in their daily lives as well as in their learning. Generally, less shy students do better in their self-expression. This difference is caused by extrovert students’ tendency to express themselves more frequently and to be more self-confident than others. However, speaking activities in a virtual world work on shy students’ self-expression. Second Life seems to help students to overcome their shyness that keeps them from trying to express themselves. This implies that students with low self-expression can overcome their weakness in a virtual world and can improve their self-expression through activities in a virtual world.

What the results of this study imply is that media can be a useful environment for students with certain characteristics. Further studies are needed to see if a virtual world can be a curing environment for shy students’ depression, low self-concept, social and emotional adjustment disorder, and low self-efficacy.
References


Harnessing Collaborative Annotations on Online Formative Assessments

Jian-Wei Lin1* and Yuan-Cheng Lai2

1Department of International Business, Ching Yun University, Taoyuan, Taiwan // 2Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan // jwlin@cyu.edu.tw // laiyc@cs.ntust.edu.tw

*Corresponding author

(Submitted July 07, 2011; Revised October 05, 2011; Accepted October 31, 2011)

ABSTRACT

This paper harnesses collaborative annotations by students as learning feedback on online formative assessments to improve the learning achievements of students. Through the developed Web platform, students can conduct formative assessments, collaboratively annotate, and review historical records in a convenient way, while teachers can generate formative assessments, manage collaborative annotations, and spot blockades to student learning in an efficient way. The experiment spanned one semester and underwent midterm and final exams. The results showed that adopting collaborative annotations on formative assessments significantly enhanced the learning achievements of students and extended retention time in the proposed system. Through a questionnaire and interviews, most students expressed positive attitudes toward using the proposed system.

Keywords

Online formative assessment, Collaborative annotation

Introduction

Formative assessments, continuously embedded in the teaching and learning process of a curriculum, attempt to improve learning achievements by offering feedback in the process. Summative assessments are used to check learning achievements at the end of the curriculum (Bell & Cowei, 2001). The point when the assessment occurs in a curriculum distinguishes these two types of assessments. The objective of formative assessments is to provide feedback to students rather than to evaluate them for course grades. Such feedback used in educational contexts is generally regarded as crucial for improving learner knowledge and skill acquisition (Pridemore & Klein, 1991). The feedback also gives learners the opportunity to develop their cognitive strategies and to rectify misconceptions through the training period (Azevedo & Bernard, 1995).

After teaching the course “Electrical Commerce” numerous times, it became clear to the author that students prepared for midterm or final exams at the last minute instead of ongoing preparation, leading to the poor preparation of students to take exams and the inadequate ability to evaluate their own knowledge about course material. Moderate adoption of formative assessments in the course drives students to practice course material in their daily lives. However, even given an adequate online formative assessment system, low participation rate among students significantly influences learning effects and in turn downgrades learning achievements. Costal (2010) indicated low participation rate as a critical concern to an online formative assessment in tertiary education. Henly (2003) reported that low ranking students, who are most likely to benefit from using an e-Learning system, were less likely to fully use online resources. Similarly, in the research of Buchanan (2000), only one tenth of undergraduate students had ever used an online formative assessment tool during a semester. Motivating students to use online formative assessment tools is an essential issue and designing appropriate methods to motivate students to complete formative assessment tasks is an important direction for future research in this area (Costal et al., 2010).

Web technology has proven that online collaborative annotations can improve both learning achievements and motivation. However, previous works have mainly focused on collaboratively annotating electronic materials or documents (Su et al., 2010; Yang et al., 2004; Yeh & Lo, 2009). Works discussing the learning performance and behavior of students when harnessing collaborative annotations in the process of online formative assessments is scarce.

Sustaining continuous formative assessment activities in a curriculum in a traditional learning environment is also time-consuming for a teacher. This is because teachers must prepare test content, evaluate student performance, identify student- learning barriers, and provide appropriate feedback. Teachers often encounter many students...
simultaneously and face pressure to adhere to scheduled instructional progress; therefore, it becomes more difficult for teachers to effectively administer formative assessments and offer appropriate feedback (Bransford et al., 2000; Wang, 2008). An efficient online formative assessment could help teachers promote teaching quality and student learning efficiency.

This study first uses collaborative multimedia annotations as feedback for online formative assessments, to raise student participation rate and to enhance student summative assessments. Collaborative annotations are used to scaffold learning to provide the necessary information to correctly answer particular questions. The proposed system can then facilitate both student learning and teacher management in formative assessments. Using the developed Web platform, students can conveniently conduct formative assessments, collaboratively annotate, and review historical records, and teachers can efficiently manage formative assessment activities, and instantly spot blockades to student learning to adjust teaching strategies. The system is currently available at the web address: http://ec1.it.cyu.edu.tw/AppCertForEC/Student/user_login.aspx in the Chinese version.

The experiment lasted for one semester and underwent midterm and final exams. The results showed that adopting collaborative annotations on formative assessments significantly enhanced student learning achievement and substantially extended student retention time of the proposed system. Most students expressed positive attitudes toward using the proposed system through questionnaires and interviews.

The rest of this paper is organized as follows: Section II introduces related literatures. Section III elaborates the proposed approach and system. Section IV presents the experimental results. The conclusion is given in section V.

Related works

Black and William (2009) suggested that providing feedback is essential in formative assessments to advance learners forward. Students receiving feedback may take steps to remedy whatever weaknesses the assessment exposes (Buchanan, 2000). Maughan (2001) stressed that the feedback of on-line formative assessments provides information that can instantly respond to the learning and teaching process to highlight needed improvements. Scant works have adopted learning feedback on formative assessments (Buchanan, 2000; Henly, 2003; Wang, 2008), and have simply focused on providing predetermined answers as learning feedback. For example, Henly (2003) used a commercial web-based formative assessment system, WebCT (http://www.webct.com), for dental education, in which students can access correct answers immediately after finishing their assessments. Buchanan (2000) developed a web-based formative assessment system, PsyCAL, for a psychology course. An important feature is that for each multiple-choice question answered incorrectly, the system provides a reference (appropriate explanation sections of textbooks) for students to further understand the reasons, instead of directly providing the correct answer. Scholars view feedback interaction as an essential component of formative assessment (Wang, 2007). Pachler et al. (2010) explicitly indicated focal points in the development of on-line formative assessment tools as 1) functions of feedback and 2) sharing outputs and ideas with peers. Costal (2010) stated that providing a direct resource increases the possibility of students assessing online formative assessment. Thus, providing appropriate learning feedback on formative assessments has become a major concern.

Annotating refer to marks readers make on reading materials. Annotations a reader makes will be helpful to subsequent readers (Hwang et al., 2007). Adopting collaborative annotations on e-materials and e-documents has proven to enhance not only learner knowledge, but also their motivation (Hwang et al., 2007; Su, et al., 2010; Yang et al., 2004; Yeh & Lo, 2009). For example, Hwang et al. (2007) proposed a web-based annotation system, capable of collaboratively sharing multimedia annotations on learning materials. Yang et al. (2004) developed a system to provide annotation creation and real-time discussion for collaborative learning. Su et al. (2010) also designed an annotation management system for sharing collaborative annotations on learning documents. However, studies discussing the learning performance and behavior of students when harnessing collaborative feedback in the process of an on-line formative assessment are scant. Certain researchers have admonished, “formative assessment effectiveness across samples and educational contexts is not well known” (Wilson et al., 2011).

This study proposed an online formative assessment tool for the course “Electrical Commerce,” a mandatory course in major and specialist programs, to enhance student willingness to discuss, refer, and rectify their misconceptions on
quizzes of formative assessments, by using collaborative annotations. Annotations and knowledge from peers serve as scaffolding feedback for students, to not only motivate learner interest, but to also reduce frustration (Shute, 2008).

Although most works emphasize a required learning mechanism in formative assessments for students, an efficient management mechanism is also important for teachers to administer formative assessments. Chen and Chen (2009) presented a mobile formative assessment system to help teachers precisely assess individual learning performance, but they did not provide learning feedback for students. Lazarinis (2010) presented an adaptive web testing system for teachers to manage and use different adaptive rules to create formative assessments for students. Testing procedure adaptation relies on performance, prior knowledge, and the goals and preferences of test participants.

This study not only adopts collaborative multimedia annotations as feedback for online formative assessments, but also develops the proposed system to facilitate both student learning and teacher management of formative assessments.

**Proposed approach and system**

**The proposed approach**

The formative assessment quizzes are based on textbook material and handouts. To build the Question Bank Database for formative assessments, teachers have to first build representative questions and their corresponding answers and a question bank for each chapter. This study uses multiple-choice format for most questions.

In the proposed approach (Figure 1), the operation procedure consists of two phases: Test and review phases. Students in the test phase take an on-line formative assessment, which pertains to one chapter of a curriculum. Following the test phase, student learning records are stored in the User Learning Portfolio Database. The procedure then enters the review phase, consisting of the following steps:

- **Step 1:** After finishing the test phase, students obtain their test results, including their own answers, the correct answers, and the question restatements, to allow them to monitor their knowledge of the subject. A student is supposed to check every incorrectly answered question (IAQ). Comparing the correct answers with the student answers detects these IAQs. The review process in the assessment terminates after the student addresses every IAQ. Otherwise, the student must continue to examine the next IAQ.

- **Step 2:** When encountering an IAQ, a student reviews it and its corresponding multimedia annotations extracted from the Collaboration Annotation Bank Database. These annotations uploaded by peers are stored in this database.

**Figure 1. Flowchart for the operation procedure**
Step 3: If the shared annotations are meaningless, the student is encouraged to figure out the whys through surfing the Internet or textbooks, and then annotate it for sharing between peers. To boost student desires to participate, several measures have been taken as incentives. First, a Credit Bulletin Board presents to rank the number of given credit each author has gained. Second, a Contribution Bulletin Board presents to rank the number of each author’s annotations. Then the student goes back to Step 1.

Step 4: If the shared annotations are beneficial, the student directly returns to Step 1 without making any annotation.

Every formative assessment follows the same procedure above. Even after a period, students can still login to review their previous (or historical) formative assessments and the review process follows the review phase (Figure 1).

Providing students with the support of collaborative annotations may facilitate acquiring comprehensive concepts, sustaining connections to the online learning community, and obtaining learning skills to cope with academic problems (Chen & Chao, 2008). Therefore, using our proposed approach, students can take advantage of this cooperative community through the following use scenarios.

Scenario 1: When reviewing the IAQs sequentially, a student finds if there are existing beneficial annotations. If yes, he/she comprehends the whys and further rectifies his/her misconception. If no, he/she figures out the whys by himself/herself through surfing the Internet or textbooks, and then annotates it by copying (and pasting) the Internet address or inputting the textbook location (what page and paragraph). Later, he/she will receive a notification e-mail from the system, which contains a hyperlink linking to the newly inputted annotation.

Scenario 2: A student who does not login to the system for several days will receive an email reminder, which notifies the user to use this system more often. When connecting to the system, the student can observe whether his ranks in the Contribution Bulletin Board and the Credit Bulletin Board are low, thus alerting him that his classmates studied harder than he did. This scenario may also raise his interest to keep up with mainstream studying in the class.

Scenario 1 exhibits how a student contributes and benefits to and from the learning community. With these supports, students realize useful resources and develop their understanding through exploring resources. Scenario 2 demonstrates how such a community raises student awareness of learning progress, which may enable a sustained connection between community members.

Teachers serve as moderators in such an environment to encourage and supervise annotating activities because a lack of moderator feedback would reduce motivation to participate in activities (Chao & Chen, 2009). To prevent inappropriate annotations, the teacher periodically and regularly patrols if the newly inputted annotations are beneficial and then deletes or modifies them if necessary, using the back-end administration platform. The teacher also informs a student of his/her improper annotations through the system so that he/she can modify them online. Students can also report inappropriate annotations to the teacher for further handling. Besides, the questions with high failure rate, annotation quantity, or review time, can be deemed as the challenging or puzzling questions for most students. After analyzing the portfolio of these questions, a teacher can offer guidance by uploading recommendatory annotations and teaching related concepts in class to clarify student misunderstandings. A teacher understands the learning bottlenecks most students face, and further intervenes to offer necessary help. Through the above circumstances, providing an adequate management platform for formative assessments can substantially enhance teaching quality.

The developed system

The structure of the developed system mainly comprises two major parts (Figure 2): (1) Access Database and (2) Web Application, built on Window Server 2003 Internet Information Server (IIS) and written by Microsoft ASP.NET, JAVASCRIPT, and HTML. The Access Database is comprised of a number of relational tables to store the necessary data.

The web application is divided into two parts: the front-end User Platform and the back-end Administrator Platform. The front-end platform is available for users to perform activities on formative assessments, whereas the back-end platform is only available for administrators (teachers) to manage user activities. The back-end Administration
Platform consists of the main management functions: Formative assessment management (FAM), collaborative annotation management (CAM), and other miscellaneous functions including summative assessment management (SAM), remote user management (RUM), and statistic analysis (SA). The striking manipulation and the corresponding GUIs are presented as follows.

**FAM**

The FAM generates web pages for the test content of the formative assessment and identifies student performance and learning barriers (the questions that most students failed). Figure 3 demonstrates some snapshots in which a teacher configures a formative assessment. The upper-left window (Step1) sets the scopes of a formative assessment and the upper-right window (Step 2) shows the setting of its name, designated class, start time, and end time (Step 2). The bottom window shows the corresponding test content of the formative assessment, which students are going to take it. All students take the same quiz, but the questions are arranged randomly to prevent cheating.
CAM

The CAM allows a teacher to administer collaborative annotation activities, such as modifying or deleting improper annotations. After finishing a formative assessment, a student can review the results, input and view annotations, receive a notification e-mail when there is a newly inputted annotation, or review historical records. Figure 4 demonstrates some snapshots. The upper-left window shows the historical list of formative assessments that a student has taken, whereas the upper-right window shows the results of a formative assessment. The bottom window shows the composition of a new annotation for a question using the text editor, embedded by a free software component FCKEditor (http://www.fckeditor.net). Using the Edition Tool Set, users can upload multimedia annotations by inputting text, images, tables, flash objects, and hyperlinks in the Edition Area and further arrange the layout.

![Figure 4. Snapshots of viewing results and editing multimedia annotations](image1)

![Figure 5. Example of collaborative annotations corresponding to a question](image2)
Figure 5 shows the annotations that contribute to a question, where the author list (Contribution Bulletin Board) is shown in descending order according to the number of annotations. Students who are industrious to generate the most annotations are placed at the top of the author list. Users can select a desired author or all authors to read his/her or all annotations, respectively, which are displayed at the bottom. If users feel that an annotation is beneficial, they can give credit to the author. Users can notify the teacher of any improper annotation for further handling.

**Miscellaneous functions**

Further miscellaneous functions include SAM, RUM, and SA. The SAM is similar to FAM, but for summative assessments. The content refers to that of formative assessments, but has slight modifications. Each student receives the same summative assessment questions, but in random order. The RUM is in charge of managing the user accounts, whereas the SA is in charge of computing the statistics (grades) of the formative and summative assessments.

**Educational evaluation**

**Objectives**

This evaluation compares the proposed system – the Collaborative Annotation Formative Assessment System (henceforth referred as CAFAS) – with the system without collaborative annotations – the Non-Collaborative Annotation Formative Assessment System (henceforth referred as NCAFAS). Upon completion of a formative assessment, both systems give students feedback about their performance, including questions, their answers, the correct answers, and scores. However, with CAFAS, students are able to input and review their collaborative annotations, but NCAFAS students have no such support. To conduct the evaluation, the NCAFAS must be developed. The modular design of the developed CAFAS simplifies building the NCAFAS with only slight modifications to the inner software structure of CAFAS. Building the NCAFAS requires simply removing the related modules, for example, the CAM module in the Administration Platform and the CA activity module in the User Platform in Figure 2.

This study was administered to three 3rd-year classes of the Department of International Business in Ching Yun University, with an approximate enrollment of 180 students. The first class, consisting of 53 students, was assigned as the experiment class and used the CAFAS. The second class, comprising 60 students, was assigned as the control class and used the NCAFAS. The third class, consisting of 51 students, was only assigned to validate and analyze the content reliability of the pretest and posttest.

The current study addresses the following issues: (1) Whether CAFAS, compared to NCAFAS, enhances learning achievements and motivation; and (2) Analyzes the behaviors of annotating and reviewing within the experiment class. Finally, the questionnaire and interview were conducted in the experiment class.

**Research tools**

The “Electronic Commerce” course includes seven chapters: “Basic Network Introduction,” “Basic Electronic Commerce Introduction,” “Basic Electronic Business Introduction,” “The Strategy and Method of the Electronic Business System,” “The Plan and Design of the Electronic Business System,” “The Implementation of the Electronic Business System,” and “The Maintenance of the Electronic Business System.” These chapters were taught in both classes by face-to-face lectures through textbook and PowerPoint material. The test content of any formative assessment originates from teaching materials, ranging from those that focus on knowledge and comprehension to those that focus on application and analysis. To establish substantial question banks, two teachers cooperated to collect and edit the question banks, generating 66, 66, 125, 119, 85, 90, and 50 multiple-choice questions for Chapters 1 to 7. A typical question may address a specific term, for example, “What is the phenomenon of the Long Tail,” followed by four possible options, with only one correct option.

Formative assessment quizzes directly reflect summative assessments of the course. Thus, students using the system become familiar with the type of questions they are required to answer on their midterm and final exams. Formative
and summative assessments have few overlaps. Even though similar types of questions appear on both assessments, the questions in a summative assessment are not directly copied from formative assessments, thus eliminating the problem of students memorizing questions and answers. The midterm test content ranges from Chapters 1 to 4, and that of the final ranges from Chapters 1 to 7. The two classes received the same test content.

Formative assessments mainly serve to motivate students to extensive and deep study and to further engage them in annotating in the experiment class. Therefore, this work does not report the test scores in all formative assessments in the following sections.

Research procedure

This research adopted a quasi-experimental design method. Both classes were taught how to use their designated systems at the onset. The experimental treatment lasted for one semester for two hours each week. Both classes adhered to the same schedule. Upon completion of each chapter, both classes received the same formative assessment of the chapter. After finishing one formative assessment, both classes reviewed their results of the assessment. However, students in the experimental class performed annotating activities to fulfill remedial learning. The two classes were encouraged to not only review the learning materials, but also their historical formative assessments before the summative assessments. The midterm and final, taken during the middle and end of the semester, were administered as summative assessments to individually evaluate the learning achievements of the two classes.

To verify possible significant difference in the prior knowledge of students in both classes, a study regarding prior knowledge was conducted before the evaluation. To assure pretest validity and reliability, 2 experts reviewed the content of the pretest, which was then tested by 51 students in the third class. After that, inappropriate questions were removed according to the corresponding difficulty and discrimination levels, resulting in 38 multiple-choice questions and Cronbach’s α being 0.83.

Table 1 shows the results of the independent samples t-test, showing no significant difference in the average scores of the basic learning backgrounds between the experiment and control classes (\( t = -0.62, p > .05 \)). The basic backgrounds of the students in the both classes showed no significant difference.

<table>
<thead>
<tr>
<th>Classes</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>53</td>
<td>36.11</td>
<td>13.13</td>
<td>-0.62</td>
<td>0.53</td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>37.81</td>
<td>12.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The validity and reliability analysis of both summative assessments were handled the same as that of the pretest, resulting in 62 and 82 questions, with Cronbach α at 0.86 and 0.88, respectively.

Data collection and analysis

To analyze participant preference, both systems recorded participant activities as logged data, including login time, activity types (reviewing, annotating, or testing), source IP (locations, such as home or school), and stay period (the time a visitor spends on the system). The SPSS was used to conduct the statistical analysis.

Comparison between the experiment class and control class

The independent samples t-test was chosen to determine if there was significant difference in the different measures between the experiment class and the control class. The results are shown in Table 2.

The mean score of the midterm of the experiment class (mean = 61.74) is significantly higher than that of the control class (mean = 52.48) (\( t = -3.50, p < .05 \)). Similarly, the mean score of the final of the experiment class (mean = 55.55) is also significantly higher than that of the control class (mean = 47.93) (\( t = -3.06, p < .05 \)). These two identical
results unveil that the use of CAFAS facilitates peer interaction in a collaborative learning context, allowing students in the experiment class to obtain shared annotations and to brainstorm. Therefore, they obtained deeper understanding of online questions in formative assessments and acquired enhanced learning achievements.

The mean of the review time before the midterm of the experiment class (mean = 104.94) is significantly higher than that of the control class (mean = 10.92) \( (t = -6.74, p < .05) \). Similarly, the mean of the review time between the midterm and the final of the experiment class (mean = 180.43) is also significantly higher than that of the control class (mean = 46.77) \( (t = -3.65, p < .05) \). These two identical results also reveal that students in the experiment class obtained more aids and resources from the collaborative annotations. Compared to students in the control class, students in the experiment class spent more time annotating, discussing, and sharing their opinions on a formative assessment.

Table 2. Results of the independent samples \( t \)-test

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>52.48</td>
<td>14.98</td>
<td>-3.50</td>
<td>.001</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>61.74</td>
<td>12.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Exam Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>47.93</td>
<td>11.55</td>
<td>-3.06</td>
<td>.003</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>55.55</td>
<td>14.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review time before the Midterm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>10.92</td>
<td>43.00</td>
<td>-6.74</td>
<td>.000</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>104.94</td>
<td>93.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review time between the Midterm and the Final</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>60</td>
<td>46.77</td>
<td>184.80</td>
<td>-3.65</td>
<td>.000</td>
</tr>
<tr>
<td>Experiment</td>
<td>53</td>
<td>180.43</td>
<td>201.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison within the experimental class

The paired-samples \( t \)-test was chosen to determine if there was significant difference in the different measures within the experiment class. The results are shown in Table 3.

For annotation quantity, the mean between the midterm and the final (mean = 7.42) is significantly higher than that before the midterm (mean = 3.98) \( (t = -3.37, p < .05) \). Similarly, the mean of review time between the midterm and the final (mean = 180.43) is also significantly higher than that before the midterm (mean = 104.77) \( (t = -3.74, p < .05) \). This is because the students were unfamiliar with the CAFAS at the beginning and required a period to become acquainted with it. With the passing of time, students learn how to share collaborative annotations and receive more benefits from them. As the number of annotations increases, more substantial annotations are generated to meet student requirements. This study originally thought that learning achievement on the final should be obviously higher than that of the midterm. However, oppositely, the mean score of the midterm (mean = 61.74) is significantly higher than that of the final exam (mean = 55.55) \( (t = 3.23, p < .05) \). This strange phenomenon may be due to the larger scope of the final. Thus, it is much more difficult for students to prepare for the final exam.

Table 3. Results of paired-samples \( t \)-test

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Pair of Annotation Quantity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before the Midterm</td>
<td>3.98</td>
<td>53</td>
<td>8.52</td>
<td>-3.37</td>
<td>.001</td>
</tr>
<tr>
<td>Between the Midterm and the Final</td>
<td>7.42</td>
<td>53</td>
<td>12.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Pair of Review time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before the Midterm</td>
<td>104.94</td>
<td>53</td>
<td>93.11</td>
<td>-3.74</td>
<td>.000</td>
</tr>
<tr>
<td>Between the Midterm and the Final</td>
<td>180.43</td>
<td>53</td>
<td>201.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Pair of Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midterm Score</td>
<td>61.74</td>
<td>53</td>
<td>12.84</td>
<td>3.23</td>
<td>.002</td>
</tr>
<tr>
<td>Final Exam Score</td>
<td>55.55</td>
<td>53</td>
<td>14.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A Pearson Correlation was employed to analyze the correlation between the final score, total annotation quantity, and total review time within the experiment class. A significant positive correlation also exhibited between the total annotation quantity and the total review time (the correlation = 0.575, \( p < .01 \)), and between the total review time and the final score (the correlation = 0.429, \( p < .01 \)). Finally, a significant positive correlation demonstrated between the final score and the total annotation quantity (the correlation = 0.337, \( p < .05 \)). Thus, these three variables have a
moderate positive correlation with each other. Therefore, it is reasonable to infer that more annotations or more review time is conducive to better learning achievement.

**Questionnaire and interview**

To understand student satisfaction in terms of their concerns, a questionnaire with a Likert scale ranging from 5 (*strongly agree*) to 1 (*strongly disagree*) was given to the experiment class at the end of the study. This questionnaire was based on Su et al. (2010) and further modified to evaluate student feelings about using the CAFAS. Among 53 students in the experiment class, 51 valid questionnaires were collected and used for the data analysis. After completing the questionnaire survey, short interviews were conducted for 9 selected students to elicit their actual feelings.

<table>
<thead>
<tr>
<th>No</th>
<th>Question</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Does the CAFAS provide a convenient environment (friendly user interfaces)?</td>
<td>4.21</td>
<td>0.73</td>
</tr>
<tr>
<td>2.</td>
<td>Does the CAFAS have good stability (quick response)?</td>
<td>3.75</td>
<td>1.11</td>
</tr>
<tr>
<td>3.</td>
<td>What is your overall satisfaction degree toward the CAFAS?</td>
<td>4.00</td>
<td>0.89</td>
</tr>
<tr>
<td>4.</td>
<td>Does the CAFAS aid you in your learning in Electronic Commerce?</td>
<td>3.85</td>
<td>0.91</td>
</tr>
<tr>
<td>5.</td>
<td>Does the CAFAS stimulate you to spend more time on the system?</td>
<td>3.86</td>
<td>1.01</td>
</tr>
<tr>
<td>6.</td>
<td>Does annotating impose an additional load?</td>
<td>3.95</td>
<td>1.04</td>
</tr>
<tr>
<td>7.</td>
<td>Do you look into the questions that have the highest review time?</td>
<td>3.69</td>
<td>1.03</td>
</tr>
<tr>
<td>8.</td>
<td>Do you look into the questions that have the most annotations?</td>
<td>3.79</td>
<td>1.05</td>
</tr>
</tbody>
</table>

The questionnaire results, shown in Table 4, reveal positive feedback for most of the evaluated aspects. The results of questions 1 to 3 show that most students were satisfied with the CAFAS and agreed that it is a stable and convenient system. For example, one interviewee stated, “The user interfaces are friendly. It is easy to use even for the first time.” Conversely, there was also one opposite opinion: “Sometimes the system responded slowly. However, it may relate to the low network bandwidth of my home PC.”

The results of Question 4 show that most students also agreed that the CAFAS is a practical auxiliary tool for learning Electronic Commerce. Most students generally agreed that the support of shared annotations was useful in providing different aspects of knowledge. For example, one interviewee stated, “Some recommended annotations and key sentences provided by other peers actually helped me to better understand the questions. Sometimes I pondered the meaning of the annotated sentences and content.” One interviewee stated, “I did not have to review the questions that I understood before the exams so that I could focus on reviewing incorrectly-answered questions on my historical formative assessments.” Conversely, 3 interviewees simultaneously expressed that they preferred reviewing the materials through papers or textbooks rather than using computers before summative assessments. One of the 3 devised an approach to cope with such a dilemma, saying, “Before the exams, I copied the content of my historical formative assessments and pasted it into Microsoft Word, and then attached useful annotations to the corresponding questions. After proper arrangement, I printed them so that I could read them in my own way.”

The results of Question 5 show that the system moderately stimulated students to spend more time on it. Six of 9 interviewees expressed that they often reviewed or annotated at home. One stated, “I spent more time on the system, especially before the day of summative assessments because it provided sufficient question banks and thus became my important resource to practice.” One stated, “I enjoy praise when someone gives credit to my annotations and this inspires me to contribute more.” Another interviewee commented, “The collaborative annotations interest me because they can present in multimedia form, which is much easier to understand.”

Certain students expressed that they prefer independent study, rather than collaboration, whereas others expressed that they prefer their peers as a source of guidance. Some even asserted that they dislike sharing their own knowledge with others. A few students stated that they were not enthusiastic about participating in collaborative discussion. Accordingly, we speculated that learning styles (Huang et al., 2009), including independent, dependent, competitive, collaborative, avoidant, and participant, may also play a crucial factor in annotating and reviewing behaviors in formative assessments and such speculation needs further verification in the future.
When investigating whether annotating imposes an additional load (Question 6), 16 of 53 students (30%) agreed; 28 were neutral (53%), and the other disagreed (17%). Although the students welcomed collaborative annotations, some suffered from an extra load associated with annotating. This may be because (1) some felt there were too many formative assessments; and (2) some felt frustrated with numerous incorrectly answered questions in the formative assessment. One interviewee expressed, “At the beginning, I was interested in a collaborative environment. However, after several months, I felt there were too many questions needing to be addressed and eventually I grew tired of annotating.” Another interviewee stressed, “I was certainly frustrated by too many incorrectly answered questions and then I reluctantly annotated or even skipped them.”

From the results of questions 7 and 8 in the questionnaire, we infer that questions with more review time or more annotations are the topics of most concern to students or that confuse many students. A teacher can see the corresponding annotations of students and further assess their level of comprehension for instructional intervention. Harnessing collaborative annotations on online formative assessments are an effective way for teachers to identify most students’ hurdles. Thus, the teacher can spend more time explaining these points to further clarify most student misunderstandings. Questions with more annotations may encourage students to look into and discuss more details. One interviewee remarked, “I was curious about a question that had been annotated by many peers, because I wanted to know their arguments. I also added my viewpoints if necessary.” Another interviewee stated, “I looked into a question with many annotations for more cues when I did not understand its meaning or its technical terms.”

**Conclusion**

This research adopted collaborative annotating on formative assessments to boost learning efficiency. The Administration Platform of the developed system CAFAS enables a teacher to organize formative assessments and to manage annotation activities, whereas the User Platform allows students to conduct and review formative assessments and to annotate collaboratively.

The comparison between the experiment class (with CAFAS) and the control class (with NCAFAS) shows that the mean scores of the midterm and final of the experiment class are both significantly higher than those of the control class. These results reveal that using the CAFAS enhances learning achievement. This finding is consistent with that of Hwang et al. (2007), who stated that the use of shared annotation on e-material enhances learning achievement. In addition, review time before the midterm and between the midterm and final in the experiment class are both significantly higher than those of the control class are. These results show that the use of the CAFAS moderately enhances student motivation to spend more time on the course.

Within the experiment class, the means of annotation quantity and review time between the midterm and the final are both significantly higher than those before the midterm. These results show that students are willing to devote more annotations and consider the CAFAS as an important learning resource when they gradually acquire more substantial aids through collaborative annotations. The results of correlation analysis exhibit three variables: annotation quantity, review time, and the final score have positive correlations.

The results of the questionnaire and interview reveal that most students have positive attitudes toward the CAFAS. They agree that the CAFAS is a convenient and stable environment. Such a system moderately stimulates students to spend more time on it and aids their learning in Electronic Commerce. The questions with much review time or many annotations can serve as diagnostic cues for a teacher in exploring students’ learning context and in turn adjusting instruction.

To establish such a system, teachers must first pay some efforts on generating multiple-choice questions at the beginning. However, an established quiz database can persistently afford current and prospective students the opportunity to benefit from the system. The current system only provides the user interface for teachers to manually input questions. Future works should develop the helpful function of generating multiple-choice questions and choosing distracters by applying Article Intelligence (AI) techniques. Thus, the burden of teachers in generating questions and distracters can significantly reduce.
References


Effects of Game Design Patterns on Basic Life Support Training Content

Sebastian Kelle¹*, Roland Klemke² and Marcus Specht²

¹Stuttgart Media University, Nobelstraße 10, 70569 Stuttgart, Germany // ²Open University of The Netherlands, Centre for Learning Sciences and Technologies, Valkenburgerweg 177, PO box 2960, 6401DL Heerlen, The Netherlands // ske@ou.nl // rkc@ou.nl // spe@ou.nl

*Corresponding author

(Submitted July 14, 2011; Revised September 20, 2011; Accepted November 06, 2011)

ABSTRACT

Based on a previous analysis of game design patterns and related effects in an educational scenario, the following paper presents an experimental study. In the study a course for Basic Life Support training has been evaluated and two game design patterns have been applied to the course. The hypotheses evaluated in this paper relate to game design patterns that have been used for learning functions, expected to enhance the learning outcome and user experience. An experimental design has been carried out in order to get insight about effects of individual and combined game patterns in a Basic Life Support course. Based on the according educational objectives, the effects of two different game design patterns relevant for learning (a timer pattern and a score pattern) have been evaluated. This game was prototypically developed targeting the application on the healthcare domain (basic life support). The results show a significant interaction effect of the two patterns on the learning gain, as well as a strong covariate influence of the learners’ age.

Keywords

First aid, Learning, Game-based learning, Game design patterns, Serious games

Introduction

The design of educational or serious games is a very complex process. Two antagonistic principles have to be united: the achievement of educational objectives (serious aspect) and meaningful gameplay (game aspect). Indeed, there are different instructional design approaches that help building the bridge between these two aspects. This can be achieved with the employment of pedagogical methods that overall help learner motivation, while adapting to the different requirements of a multitude of learning contexts, as will be described in the next section. However, on the more technical end of the scale, building learning games also to a large extent requires detailed technical modeling and implementation, a challenge that touches upon technical standards, as pointed out in a previous paper (Kelle et al., 2011).

One of the possibilities to structure and simplify the quest of “how to design learning games properly” is the principle of using design patterns, which have been described for instance by Björk and Holopainen (2004). This approach bears a structured and expedited approach to the design and creation of digital games. Game-patterns encapsulate common design problems and solutions for those and game designers typically combine several patterns for good gameplay. One of the main principles of game design patterns is the fact that they modulate and instantiate each other; for example, the pattern of “rewards and penalties” automatically demands the existence of some sort of “score” or “resources.” Patterns can also be in conflict with each other, such as “real-time-play” and “turn-based-play.” For this reason, in a game design, there generally shouldn’t be conflicting patterns, because, for example, if a game is turn-based the fluent characteristic of a real-time game might be lost. For evaluation of a game design, in the end, play-testing (Schell, 2008) is necessary, which yields information on the user experience. The limitation here is that end-user testing brings the risk of not giving insight on what patterns exactly have contributed mainly to the success or failure of the game. In the experimental context, it is therefore essential to alternate between certain combinations of patterns to isolate what makes the game work and what not. Hence, for iterative game design incorporating end-user testing, this approach also bears high potential.

This paper describes an experiment based on previous research in which we have analyzed pattern-based approaches in the field of commercial game design (Björk & Holopainen, 2004), provided a mapping onto learning functions and educational objectives, and evaluated the mapping with experts in Technology Enhanced Learning. While the method of developing games with the help of game design patterns is common sense, in the field of education, evidence of the efficiency of such pattern-based serious games is scarce. In the following section the main existing...
findings will be discussed. After that we describe an experiment that applies two selected game patterns in a learning game and evaluate their effects on knowledge gain and user experience.

Related work

As mentioned, literature evidence for use of patterns in serious gaming is not very numerous; however, in addition to our main inspiration source for the idea of educational game design patterns (Björk and Holopainen, 2004), some other relevant leads exist, which together cover a relatively broad range of learning contexts. We will briefly give an overview of what has been reported so far by the research community.

Gunter et al. (2006) combine educational theories with a model for the design process they call RETAIN (Relevance Engagement Translation Assimilation Immersion Naturalization), which they base on well-established theories of Gagné and Keller (Gagné, 1985; Keller, 1983). The strength of this approach is argued to lie in the employment of a sound theoretical foundation relevant in motivation psychology and instructional theory. However, although the approach is pointing into the direction of drawing conclusions for the application in a pattern-based design methodology, the implementation and evaluation remains future research. Another point to be noted is that the authors address an abstract level of learning, such as the cognitive, affective and psychomotor learning domains, without targeting an actual outcome oriented learning context, specifically.

Mor et al. (2006; n.d.) choose an experience driven approach that is closer to the technical implementation side, but limits itself to the context of secondary mathematical education. They have made first experiences with using game design patterns for learning in the Kaleidoscope Project (Kaleidoscope Project, 2011), when the objective was teaching Mathematics to young learners, deriving a more general pattern based approach for the use in Technology Enhanced Learning. These findings led to the implementation of a web-based tool that enables the creation and archiving of design patterns. An example for such a design pattern would be the “crescendo” pattern that deals with the problem of emerging discussions in a learning environment, spiraling towards a more rhetorical than the (more desired) reflective mode. Again, the limitation here lies within the limitation to the mathematical domain, but at least parts of the patterns could possibly be generalized. For most patterns, evidence of their actual use in the learning context is given in a qualitative overview: for example, the crescendo pattern has been implemented and tested by Cerulli (2006), indicating a fair level of positive impact. A detailed quantitative evaluation, however, is missing.

Shute et al. (2009) used a model driven approach for assessment based learning game design, using elements like highscore and resources patterns to build their learning games. Their target audience was K-12 education level students in Mathematics. The way they used assessments to leverage the gaming aspect in their approach was by conjoining games with “embedded” assessments that are hidden from the user. They establish the term “stealth assessment,” which they exemplify by modeling a competency-driven learning paradigm applied in the game “Oblivion.”

Denis and Jouvelot (2005) used a best practice-based anthology of game elements in order to achieve a high threshold of motivation. Their target domain was musical education. In their approach they used a game that trained solo parts and accompaniments of certain musical pieces, training chords and scales with players taking the roles of piano players and saxophonists in pairs of two (duets), controlling the musical interface by means of standard computer gaming equipment like joysticks and gamepads. A fair deal of freedom was given to the players, enabling improvisation, giving them ownership of their interactions.

Dickey (2006) looked at Massively Multiplayer Online Role-Play Games (MMORPG’s), and determined several elements that take a key role for player motivation. In his overview, he outlines the element role-playing that is responsible for a high identification factor of the player and the game character. Furthermore, a high emphasis is put on the element of narrative structure, which guides the activities of the player, being responsible for a high intrinsic motivation. As a particular element relevant for the learning context is the presence of quests that address the problem-solving aspect.

The notion of design patterns for educational games can be found more explicitly in the work of Huynh-Kim-Bang et al. (2010), in which several patterns are drawn from the analysis of 20 serious games examined by the authors. They
describe the following patterns: Serious Game, Game-Based Learning Blend, Instructive Gameplay, Time for Action/Time for Thought, Reified Knowledge, Museum, and Fun Reward.

These patterns rank on different abstraction levels: While the first three patterns address a very broad spectrum of educational gaming, the latter four target more concrete dimensions. Time is identified here as axis both relevant for a more intensive “action-based” gaming experience in which the player has to deal with tasks in a rapid way. On the other hand, this is counterbalanced with more contemplative phases, which yield time for thought and reflection. The pattern of Reified Knowledge, however, drives more into the direction of self-awareness of the user’s progress in the game, by manifesting certain virtual objects that represent goals and results of the game learning process. Finally, the Fun Reward pattern aims at game elements that trigger motivation for the user’s incentive to keep playing. Overall, the approach in this work provides a useful insight on how to create meaningful connections between learning and gaming. The authors state in the conclusion, however, that there is the lack of external validation.

Finally, Kiili puts forward another promising approach by aggregating a collection of educational game design patterns on his website (Kiili, 2011). His typology of patterns comprises several categories: Integration Patterns, Cognition Patterns, Presentation Patterns, Engagement Patterns, Social Patterns and Teaching Patterns. For each of these categories at least one pattern has been collected so far. This pattern library is open to suggestions for new patterns and as such could become an important repository for the community of educational game designers.

With the exception of the approaches of Mor et al. (2006), Huynh-Kim-Bang (2010) and Kiili (2011), in the approaches listed above the actual formalization of game design patterns is either not very concrete, or targeting a too narrow scope to be generalized, transferred and re-used (which is really the main purpose of design patterns). The main advantage of these findings, however, is that there are several leads that point into the direction that a pattern-based approach enhances design methodology with a direct positive impact on user experience and learning outcome.

Using the definition described by Mory (2004): “a process in which the factors that produce a result are themselves modified, corrected, strengthened, etc. by that result” and “a response, as one that sets such a process in motion,” the most relevant aspect is the notion of feedback for self-regulated learning, which was discussed in detail by Butler and Winne (1995). According to them, the mirroring of feedback to the learner is of high importance to affect cognitive engagement with tasks, using feedback of intermediate and total achievements in the learning process. The reason for this lies in the fact that an improvement of the learning experience and outcome can be measured positively if learners are given the possibility to monitor and gauge their own progress during their learning activity. As described by Verpoorten et al. (2010), this condition allows the learners to scrutinize and reflect about their newly acquired knowledge, a process which has the potential for a lasting learning success.

Revisiting Kiili, more considerations are raised with respect to feedback-induced reflection in learning game scenarios. Kiili (2008, 2011) eclectically argues for reflection as key principle in learning games. He proposes a methodology called “Reflection Walkthrough” that is derived from the user evaluation principle of cognitive walkthrough, in order to isolate potential strengths and weaknesses of a learning game design. The methodology gives insight on feedback mechanisms that trigger reflection, the support of double loop learning, and the potential risks of evaluation overhead and cognitive overload.

As a challenge that summarizes these aspects and motivates our research we quote the recommendation for future research stated by Mory (2004): “Continue to identify and test interactive patterns among the learner, the environment, individual internal knowledge construction, and varying types of feedback.”

Preparations and research questions

In our previous research, extending the work of Gunter et al. (2006), as well as Kiili (2008, 2011), we looked at several pedagogical theories and taxonomies (Kelle et al., 2011), which form the bridge between game design and learning goals and functions. The method employed for this “bridging” was a step-by-step algorithm that was evaluated with 10 experts in instructional technology who independently of each other mostly came to the same results for a pattern matching between educational and game design patterns. Different from Kiili’s method of
Reflective Walkthrough, we focused rather on the preparatory end of sound learning game design than post mortem evaluation.

In this expert we had asked our 10 experts to rank the matching of a choice of game design pattern with different learning functions. The results from these expert evaluation study led us to the selection of specific patterns, i.e., the so-called time limits pattern and score pattern and that these are especially well suited for the “monitoring” learning function, which enables the reflection of learning (and game-) progress to learners. It turned out that the score pattern achieved an average ranking of 4.64 (out of 5) for the learning function of “monitoring,” and the timer pattern achieved a ranking of 4.2. These patterns can be found in Björk and Holopainen’s compendium about game design patterns (Björk & Holopainen, 2004), and are described as follows:

- The time-limit pattern is described as the pattern for completing an action, reaching a goal, staying in a certain mode of play, or finishing a game session with a limit based on either game time or real time.
- The score pattern is described as the numerical representation of the player’s success in the game, often not only representing the success but also defining it.

These patterns individually also showed to have a relatively big disagreement factor compared between the experts’ ratings. We wondered about this and therefore decided that this requires further examination.

In the experiment for this paper we have used a classic model of three experimental groups and one control group which account for the different possible combinations of both time-limit and score pattern (for details see method section). Henceforth, we refer to the different treatment groups as such:

- T0 is the control group in which none of the patterns have been applied.
- T1 is the group that only has been exposed to the time-limits pattern.
- T2 is the group that only has been exposed to the score pattern.
- T3 is the group that has been exposed to both score and time-limits pattern.

As target domain the medical topic of basic life support and first aid was chosen, because the topic is relevant, indifferently of demographical factors, for the simple reason that everybody is at permanent risk to run into an emergency situation of serious gravity (either as victim, causer or bystander). We thus controlled the risk of introducing a bias of intrinsic demotivation due to possible lack of interest in the learning content. As source for the learning content we took the guidelines available on the European Resuscitation Council’s website (ERC, 2010).

The objective of the learning activity in this experiment was the training and re-activation of basic knowledge relevant for the learner’s reaction speed and quality of decisions in emergency situations that require a first-aid response. Hence, the main educational objectives beside knowledge gain and refreshment of existing knowledge were fast reaction times. The relation to the learning function “Monitoring,” which was strongest rated by the experts in our preparatory study was also considered for the choice of patterns applied. Monitoring in this case entails the reflection of progress and success, mirrored to the learner throughout the progress of the game. Therefore, the choice of game design patterns for this experiment was narrowed down to what could possibly best link to the main learning goals: improvement of reaction speed and quality of responses and creating corresponding in-game awareness thereof.

In order to cater for an elicitation of high response quality, mirroring of the user’s performance was needed. The most obvious way to do that was to display a game score during the experiment; in order to enhance reflection for motivation and self-awareness, as consequence of a self-monitoring learning function. The users could thus monitor their performance and gauge their own skill levels on the fly. The other objective of interest was fast reaction time. Here, the best matching design pattern was the time-limits pattern, implementing a game element that creates a time constraint and displays a timer to the user. In order to advance in the game successfully, the user interaction had to be performed inside that time limit (in our case, 60 seconds per game unit). While the level of realism in our serious game indeed was not the highest due to technical limitations, the time limit introduced a certain notion of stress, which according to Svenson and Maule (1993) can have an effect on decision framing (the opposite decision can be taken if under time pressure). In our case the purpose was to create a more realistic scenario as well as train the users for quick decision-taking. According to Zur and Bretznitz (1981), time pressure also can have the positive influence on a subject to take decision that is less risky than taken without time pressure.
The main objective of our experiment was to evaluate effects on knowledge gain and motivation catalyzed by the time and score patterns applied on learning content.

The research questions and hypotheses derived hereof are stated as follows:

(1) Will the knowledge gain of participants be significantly increased by the application of the timer and score design patterns?

**Hypothesis.** Knowledge gain will increase when both patterns are applied, in comparison to the application of only one pattern, or with none pattern, such that the knowledge gain of T3 will be bigger than of T2 and T1, and the knowledge gain of T0 is smallest.

(2) What is the role of age of participants, and previous knowledge related to medical, computing and computer gaming experience? What are correlation effects and covariates?

**Hypothesis.** We expect that the effects of time and score on the knowledge gain are independent of other variables like age, previous knowledge and computer gaming experience.

(3) What impact can be measured for the user experience in different groups and subsets of groups?

**Hypothesis.** The application of game design patterns have a positive impact on user experience which we monitored in further dependent variables like perceived suspense, perceived knowledge gain, enjoyment and users’ score in the game.

**Method**

In the operationalization we used two independent variables, i.e., we combined the use of the time-limit pattern and the use of a score pattern applied to the learning content. This resulted in four different treatments combining the two levels of the variables. Regarding the treatment groups, a 2x2 matrix design with 3 experimental sample groups and one control group could be formulated (cf. table 1).

<table>
<thead>
<tr>
<th>Score Display On</th>
<th>Time-Limit Display On</th>
<th>Time-Limit Display Off</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3 = ScoreTime</td>
<td>Both time limit and score pattern</td>
<td>T2 = Score Only score pattern</td>
</tr>
<tr>
<td>Score Display Off</td>
<td>T1 = Time Only time limit pattern</td>
<td>T0 = Control No game design pattern</td>
</tr>
</tbody>
</table>

As dependent variables we measured knowledge gain, user appreciation, game score, and perceived knowledge improvement. These dependent variables were measured with tests after the treatment (in the case of knowledge gain: before and after). Furthermore, we calculated the knowledge gain by using questionnaires applied before and after the treatment, making use of

- multiple-choice questions for scenarios upon encounter of a victim in traffic, indoors, outdoors, and revival scenario,
- test questions for terminology of AED (Automated external defibrillator) and CPR (Cardio-Pulmonary Resuscitation), as well as how to perform CPR

The knowledge gain was calculated as the difference of the sums of the number of correct and incorrect answers (see formula in results section). User appreciation was measured in terms of enjoyment of users rated on a Likert-scale. The game score was the actual final score the users achieved in the game, and perceived knowledge improvement was a self-assessment of confidence about the user’s knowledge, on a Likert-scale. To complete the portfolio of dependent variables, we also focused on user experience, and asked the users how suspenseful they found the game, and how well they had understood how the game works.

As control variables demographic information and previous experiences with computer games and Basic Life Support has been ascertained in the pre-questionnaire.
In total 133 subjects participated in the study. These 133 subjects formed 4 different treatment groups, randomly assigned according to the experimental design. In group T3 there were 36 subjects, in group T2 there were 38 subjects, in Group T1 there were 35 subjects and in group T0 there were 24 subjects. Overall, there were 47.4% of female participants and 52.6% male, with similar group distributions. The average age of participants was 32.87, and 62.9% had a university degree or higher education level.

Table 2. Report about descriptives of test samples

<table>
<thead>
<tr>
<th>Group</th>
<th>String</th>
<th>Age Mean</th>
<th>CompLit Mean</th>
<th>CompGameLit Mean</th>
<th>Med_knowl Mean</th>
<th>FirstAid Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Mean</td>
<td>41.52</td>
<td>3.3750</td>
<td>2.4800</td>
<td>2.68</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>25</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>14.295</td>
<td>.82423</td>
<td>1.41774</td>
<td>.988</td>
<td>1.472</td>
</tr>
<tr>
<td>Score</td>
<td>Mean</td>
<td>28.05</td>
<td>3.6842</td>
<td>3.5789</td>
<td>2.79</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>37</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>6.105</td>
<td>.87318</td>
<td>.94816</td>
<td>.811</td>
<td>1.143</td>
</tr>
<tr>
<td>ScorTime</td>
<td>Mean</td>
<td>34.36</td>
<td>3.7838</td>
<td>2.7667</td>
<td>2.57</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>36</td>
<td>37</td>
<td>30</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>10.450</td>
<td>.94678</td>
<td>1.22287</td>
<td>.987</td>
<td>1.175</td>
</tr>
<tr>
<td>Time</td>
<td>Mean</td>
<td>30.23</td>
<td>3.7143</td>
<td>3.2857</td>
<td>2.37</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>9.726</td>
<td>.75035</td>
<td>1.04520</td>
<td>.843</td>
<td>1.207</td>
</tr>
<tr>
<td>Total</td>
<td>Mean</td>
<td>32.86</td>
<td>3.6642</td>
<td>3.0938</td>
<td>2.60</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>133</td>
<td>134</td>
<td>128</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>11.125</td>
<td>.85790</td>
<td>1.20653</td>
<td>.908</td>
<td>1.277</td>
</tr>
</tbody>
</table>

The age distribution per group was differing in the treatment groups, due to the random assignments: Participants were older in the control group where the average age was 41 years, with the highest standard deviation. Overall there were 76 participants from Asia, 22 from America, and 35 from Europe, fairly well covering a broad range of different backgrounds. In the pre-test questionnaire, participants were also asked to give detail about their previous experience and knowledge about the topic. There was an average medical knowledge of 2.6 (out of 5), computer literacy of 3.7, and computer game literacy of 3. Most participants previously had taken none (35%) or one (37%) first-aid course (cf. table 2).

The knowledge gain was calculated as follows:

\[ K_{\text{gain}} = \sum S_{\text{post}} - \sum S_{\text{pre}} \]

with \( S_{\text{post}} \) denoting the score (\( S = 0 \) if incorrect, or \( S = 1 \) if correct) of correct answers given in the post-test, and \( S_{\text{pre}} \) denoting the score of correct answers given in the pre-test. If the sum of correct answers was higher in the post-test than in the pre-test, it meant there was a positive knowledge gain. If it would have been lower in the post-test than in the pre-test there would have been a negative number as result, which would mean that somebody knew “less” than before.

**Apparatus**

The experiment was implemented by using the Emergo (Nadolski et al., 2008) Toolkit, which is a java-based application framework and authoring environment for web-based learning games. For our aims this solution provided the right characteristics, because it was possible to create a learning game experience that has almost no distraction elements (user registration dialogs, social network feeds, etc.). In figure 1 and 2 it is shown how the two different patterns were realized in the user interface.
Results

Hypothesis 1. Knowledge gain will increase when both patterns are applied, in comparison to the application of only one pattern, or with none pattern, such that the knowledge gain of T3 will be bigger than of T2 and T1, and the knowledge gain of T0 is smallest.

While the biggest knowledge gain could be measured when both patterns were applied, the second-best learning result was achieved in the control group where there were no patterns applied. Table 3 shows the knowledge gain results.

<table>
<thead>
<tr>
<th>Score on</th>
<th>Score off</th>
</tr>
</thead>
</table>
| Time on  | KnowGain = 1.9167  
Std. Deviation = 1.79483  
KnowGain = 1.4286  
Std. Deviation = 1.57715 |
| Time off | KnowGain = .9211  
Std. Deviation =1.32301  
KnowGain = 1.6667  
Std. Deviation =1.43456 |

A univariate analysis showed significant effects when measuring between-subjects effects on knowledge gain ($F = 5.104$) at a significance of $p = 0.026$ for the combined treatment with both patterns, while the knowledge gain for both treatments with only 1 pattern or the baseline without any pattern was not significant (cf. table 4).

Table 4. Tests of Between-Subjects Effects on Knowledge Gain

<table>
<thead>
<tr>
<th>Group</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0 (control group)</td>
<td>.506</td>
<td>.478</td>
</tr>
<tr>
<td>T1 (time limit pattern applied)</td>
<td>1.924</td>
<td>.168</td>
</tr>
<tr>
<td>T2 (score pattern applied)</td>
<td>.222</td>
<td>.638</td>
</tr>
<tr>
<td>T3 (both patterns applied)</td>
<td>5.104</td>
<td>.026</td>
</tr>
</tbody>
</table>

With respect to hypothesis 1 the result showed that the hypothesis could be verified only partially. While the application of both patterns elicited the highest knowledge gain significantly, the other treatments had no significant knowledge gain; with the time limits pattern ranking second. This hints at a strong combination effect of both patterns.
**Hypothesis 2.** We expect the effects the time and score patterns on learning gain to be independent of other variables as age and previous knowledge, and computer game experience.

Here, we found that the age correlated (between subjects) significantly with knowledge gain in the groups for the treatments of both time and score patterns ($p = 0.006$), as well as with only the time pattern ($p = 0.044$), while the control and score groups did not show such a significant correlation. It is also remarkable that there was no significant correlation between previous medical knowledge as well as computer literacy of the participants and knowledge gain.

**Table 5. Analysis of Variance, using Age as covariate**

<table>
<thead>
<tr>
<th>Group / Covariate</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (covariate)</td>
<td>8.960</td>
<td>.003</td>
</tr>
<tr>
<td>T0 (control group)</td>
<td>.535</td>
<td>.466</td>
</tr>
<tr>
<td>T1 (time limit pattern applied)</td>
<td>2.928</td>
<td>.090</td>
</tr>
<tr>
<td>T2 (score pattern applied)</td>
<td>.110</td>
<td>.741</td>
</tr>
<tr>
<td>T3 (both patterns applied)</td>
<td>.619</td>
<td>.433</td>
</tr>
</tbody>
</table>

The correlation between number of times of already taken first aid courses and knowledge gain showed to be significant in the control group, thus indicating that the absence of game patterns is best for those subjects who had taken already several first aid courses. For people who had already a fair deal of computer gaming experience, a significant correlation was found in the group for the treatment with the score pattern.

Using a covariate analysis we established that there was a significant effect of age on the results, which appeared to occlude the actual effect on knowledge gain (see table 5, the effect of the age was large and highly significant with $F = 8.96$ and $p = 0.003$). Consequently, we split the test population in halves, at the median of the age of 30 (size of subgroups was slightly bigger in the group of younger participants with ratio 71/62). We then tested again for significance of the effect of the treatments on knowledge gain, for younger and older participants separately (cf. table 6). The results showed that there was no significant effect of any treatment on the knowledge gain in the set of younger participants, but the effect on the knowledge gain of older participants was significant in the subgroup that had the “time limit” treatment ($F = 6.835, p = 0.011$).

**Table 6. Analysis of variance using age split**

<table>
<thead>
<tr>
<th>Age split</th>
<th>Group</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>young</td>
<td>T0 (control group)</td>
<td>.791</td>
<td>.377</td>
</tr>
<tr>
<td></td>
<td>T1 (time limit pattern applied)</td>
<td>.291</td>
<td>.592</td>
</tr>
<tr>
<td></td>
<td>T2 (score pattern applied)</td>
<td>.045</td>
<td>.832</td>
</tr>
<tr>
<td></td>
<td>T3 (both patterns applied)</td>
<td>1.223</td>
<td>.273</td>
</tr>
<tr>
<td>old</td>
<td>T0 (control group)</td>
<td>.354</td>
<td>.554</td>
</tr>
<tr>
<td></td>
<td>T1 (time limit pattern applied)</td>
<td>6.835</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>T2 (score pattern applied)</td>
<td>.797</td>
<td>.376</td>
</tr>
<tr>
<td></td>
<td>T3 (both patterns applied)</td>
<td>1.832</td>
<td>.181</td>
</tr>
</tbody>
</table>

This result indicates that the hypothesis 2 was refuted with respect to the strong covariate influence of age of the participants. After decomposing the sample into subgroups regarding the age split of “young” ≤ 30 and “old” > 30 years it was only the time-limits pattern that showed significance in the older set of participants.

**Hypothesis 3.** The application of game design patterns have positive impact on user experience which we monitored in further dependent variables like perceived suspense, perceived knowledge gain, enjoyment and users’ score in the game.

To examine results for this hypothesis (cf. table 7), we tested effects on rather experiential dependent variables. It turned out that the effect on actual points achieved in the game was significant in the group of older participants that had received the treatment with both time and score pattern ($F = 5.411, p = 0.024$), while in all other groups and subgroups there was no significant effect on points achieved.
Another dependent variable was linked more closely to user experience: The participants were asked how much they had enjoyed playing the game. Here, no significant effect could be measured in any group (therefore not listed in Table 7). However, when asked about how much suspense they had felt during playing the game, a significant effect ($F = 7.516, p = 0.008$) could be measured in the group of younger participants that had received the treatment with both score and time patterns.

Interestingly, looking at perceived knowledge gain, in the group of older participants the treatment with the score pattern showed a significant effect ($F = 7.519, p = 0.008$) and the treatment with the time and score pattern showed a significant effect of similar value ($F = 7.065, p = 0.01$). There was no significant effect in any other variables and treatment groups, or in the subgroups of younger participants.

**Discussion and conclusion**

Looking at the main question of this research, i.e., if the combination of both patterns has positive effects on knowledge gain and user experience, the results show overall the tendency that this is the case, especially for participants of older age. On the one hand, this stresses the fact that game design patterns should not exist alone; indeed, by the very nature of their definition according to Björk and Holopainen (2004), choosing one game design pattern in most cases automatically requires the presence of other game design patterns, and so forth, inductively. The fact that we observed in our limited setup that already the presence of two game design patterns exhibited a significant combination effect on user experience and player score points towards the importance of interlinking such patterns and make them supplement each other so they provide a sound, holistic game design that suits the respective context.

With respect to the strong influence of age, it appeared at first that isolated game patterns have an even lower value than no game pattern at all when being applied to “gamify” learning content. It quickly became clear that age had a significant covariate effect that influenced the main result. It was, hence, necessary to analyze the data more in-depth, with splitting between younger and older participants at the median value of 30. This revealed that the first observation could only partially be confirmed. A significant effect on knowledge gain then only could be monitored for the treatment with the “time limit” pattern, indicating that stress induced by a timer has a positive influence on knowledge gain for older participants. An informative addition to this observation could be made when not just looking at knowledge gain, but also at other dependent variables. Indeed the actual score reached in the game and perceived knowledge gain showed significant benefit in the group of older participants who had the treatment with both score and time patterns.

It could also be observed that younger participants showed a fair deal of inertia with respect to the effect of different treatments on learning outcome. What was interesting, however, was that the treatment with both patterns in the younger group was perceived as most suspenseful. A potential interpretation of these findings could be that younger participants take more notice of the gameplay as such while not being as responsive with respect to the intended learning objective. While correlations might partially give insight to the reasons of the significance of effects on knowledge gain when looking at the whole test sample, the correlations were no longer significant after splitting between younger and older participants. This indicates that in future research quasi-randomization with equal age distributions in all treatment groups will be required.
Subsuming, our result has limitations because we only tested two patterns and the result was not significant for the younger half of the test sample. Future research in this direction should therefore try different contexts with different patterns, with particular awareness of age of the participants, and extend the fundus of data with similar or bigger sample sizes.

The theoretical background of this study, which largely foots on the paradigms of feedback and reflection, seems to withstand being put to the practical test. This can be concluded because the suggested benefit of design patterns for the gamification of learning content could be validated especially in the self-directed learning context, which is more relevant to older participants. It is, however, necessary to disclaim that our target domain of basic life support and first aid training usually is organized in a quasi-curricular fashion under the surveillance of expert instructors. The intrinsic motivation, though, to enroll for first-aid training, tends to be higher for more mature participants, as the necessity for such undertaking depends more on personal insight and experience. As such, design patterns for learning games seem to be well suited for the life-long learning context.

References


Interactive Videoconference Supported Teaching in Undergraduate Nursing: A Case Study for ECG

Ufuk Celikkan1*, Fisun Senuzun2, Dilek Sari2 and Yasar Guneri Sahin1

1Department of Software Engineering, Izmir University of Economics, 35330 Izmir, Turkey // 2Ataturk School of Health, EGE University, Izmir, Turkey // ufuk.celikkan@ieu.edu.tr // fisun.senuzun@ege.edu.tr // dilsari@gmail.com // yasar.sahin@ieu.edu.tr

*Corresponding author

(Submitted July 25, 2011; Revised September 23, 2011; Accepted November 23, 2011)

ABSTRACT

This paper describes how interactive videoconference can benefit the Electrocardiography (ECG) skills of undergraduate nursing students. We have implemented a learning system that interactively transfers the visual and practical aspects of ECG from a nursing skills lab into a classroom where the theoretical part of the course is taught. The students and the instructor in the classroom observe the activities in the skill lab in real time, while communicating with the nurse in the lab via audio and video links. An experiment was performed with the participation of 13 male and 57 female (total 70) second year nursing students—36 of who were assigned to Videoconference group (experimental group-VCG) and the other 34 were assigned to Traditional Classroom groups (control group-TCG). In the experiment, ECG knowledge levels of participants were measured by repeated tests (pretest, posttest I and posttest II) and data were analyzed with repeated measures of variances and covariance, the results demonstrating that videoconferencing contributed significantly to the improvement of ECG skills of the participants. In addition, a questionnaire was given to students along with posttest II, and the result of which indicated overwhelming satisfaction with videoconference based lecture.

Keywords

Improving classroom teaching, Videoconference in nursing, Architectures for educational technology, Cooperative/Collaborative learning, Distributed learning environments

Introduction

There are a number of reasons which may reduce the effectiveness of the clinical activities of the nursing students. These include:

• The possibility that the retention ratio of theoretical knowledge taught in the classroom may decrease gradually each day as the course progresses, reaching its lowest level immediately before clinical activities start.
• Students who are very comfortable receiving theoretical knowledge in the classroom may become excessively nervous and apprehensive in the clinic when applying the theory, never previously having seen the procedure done on a patient.
• The students are not able to ask specific questions pertaining to the clinical practice during the theory part of the course as they have no experience of the clinic, and are unaware of the potential problems to be faced in a clinical environment.

A lecture that is augmented with videoconferencing may address these problems, since the clinical activities of the nurses performed in the clinic is brought directly into the classroom during the theoretical study, providing the much needed interaction between the nurse at work and student at classroom. Since many of the hospitals and universities have a limited number of clinics, and their capacities may be insufficient to allow adequate student practice; students may lack the opportunity to closely observe an actual operation.

The use of e-learning technologies in undergraduate nursing classes and in post graduate continuing education has been accepted as a valuable teaching aid in educating trainee-nurses (Ostrow & DiMaria-Ghalili, 2005). A number of studies investigated the effectiveness of technology based delivery mechanisms in nursing education. Web-based courses, use of PDA’s, digital broadcasting, video streaming, wikis and blogs, and webcasting are several examples of technology based delivery mechanisms (Miller et al., 2005; George et al., 2010; DiMaria-Ghalili et al., 2005; Whitmore, 2005; Grassley & Bartoletti, 2009).
Delivery mechanisms can broadly be classified as synchronous and asynchronous. Synchronous delivery mechanisms involve real-time interaction between participants. Videoconferencing, (Lawson et al., 2010), webcasting, non-commercial satellite interactive television fall into this category (synchronous delivery mechanisms) while asynchronous modes of education involve the use of electronic mail (e-mail), listerv, newsgroups, web pages and bulletin boards, and also traditional methods that use books and videos. Both delivery methods have their own advantages and disadvantages. Digital broadcasting provides greater access than web-based coursework. Wikis and blogs lack real time interaction. Related expenses such as the cost of hardware, software and support staff is detrimental in the adoption of the PDA (Miller et al., 2005). Web-based courses provide time flexibility, convenience and eliminate geographical limitations in nursing education. However, the results of web-based courses on learning outcomes in terms of test scores and course grades are varied. Some studies suggest that web-based teaching and traditional classroom-based teaching cause no significant differences in grades (Lu et al., 2009; Bata-Jones & Avery, 2004). However, a major drawback of web-based courses is the lack of interactions among students and the instructor, with students demanding immediate feedback, as they do not have face-to-face communication, as in the case of traditional classrooms (Atack & Rankin, 2002). The same disadvantage is also observed when using digital broadcasting and webcasting (Ostrow & DiMaria-Ghalili, 2005). These studies support the view that interaction, which is lacking in the asynchronous mode of education, is a vital element in the learning process. The feasibility study regarding three synchronous delivery mechanisms—webcasting, internet videoconferencing and interactive satellite TV has shown that interactive satellite TV and videoconferencing creates extra requirements such as the need to allocate rooms with special connections and equipment (Jones et al., 2006). In contrast, webcasting has been found to be a suitable mechanism due to its less demanding setup requirements and lower cost.

Many different interactive e-learning technologies and methods mentioned above have been successfully implemented in different fields, such as medical, engineering, social and nursing education (Lu et al., 2010; Sahin & Balta, 2007; Monova et al., 2010; Alic et al., 2009; Ma et al., 2010). Interaction is a way of improving the quality of e-learning methods and persistence of knowledge and interactive e-learning combines the many advantages of both face-to-face education and e-learning. There is an abundance of studies on the use of e-learning in nursing education. One particular study conducted by Jang et al. focused on investigating the effectiveness of web-based teaching methods for teaching electrocardiography to nursing students (Jang et al., 2005). However, the published literature on the use of videoconferencing is very limited. In a recent study, Rush and Walsh investigated the use of videoconferencing in a diabetics clinic, in which the students in a university classroom observed and interacted with the nurse in the clinic (Rush et al., 2011). This study had a somewhat limited scope in the sense that it only measured the reactions of students to videoconference based teaching rather than measure actual learning outcomes. They recommended that the method should be applied to different areas in nursing education. Our study investigated the effectiveness of videoconferencing with an emphasis on the impact on student’s learning outcomes in a different area, namely ECG. Some other areas under consideration are cardio pulmonary resuscitation, invasive procedures and demonstrations of personal hygiene.

A course enhanced with videoconferencing has several advantages over a purely web-based course:

- Students can interact with the instructor and possibly with the patient at the clinic and obtain immediate response to questions on the practical aspects of clinical procedures.
- The instructor at the clinic receives immediate feedback from the students to enhance the teaching methods.
- The videoconferencing session can be archived for the perusal of the students who did not participate in the original course.
- Visual stimuli may encourage more perceptive and insightful student questions.

While the positive effects of technology based delivery in nursing education are undeniable, there are factors that adversely affect the adaption of technology (Washer, 2001; Mallow & Gilje, 1999). The negative factors mostly center on getting familiar with the new technology and limited resources related to time, money and equipment (Kataoka-Yahiro et al., 2011). A study done by Potempa et al. (2001) also lists similar barriers that may limit the use of technology in e-learning, which need to be taken into account when implementing e-learning based solutions.

**Interactive video conferencing and the proposed system**

Interactive videoconferencing supported learning (IVCSL) is a system used to improve the learning skills of the students studying in a traditional classroom. IVCSL aims to bring students and multiple instructors together through
a number of communication mediums, such as LAN connections, private cabling, wireless and the Internet, to provide continuous training, and to create a synergy which enhances the quality of the learning. There are basically three widely used forms of transmission in videoconferencing: Telephone (ISDN), Internet (IP) and Private Lines. ISDN, which has been in operation for at least 25 years, is usually implemented using dedicated telephone lines (Jones et al., 2006). The high cost of ISDN is paving the way for IP based videoconferencing, which provides flexible, easy to use and more cost effective transport of audio and video data (Sackett et al., 2004). Private lines are no longer the primary choice because they are inflexible and costly to maintain.

The basic elements of the IVCSL system for nursing education are classrooms, clinical skill labs, students, lecturers and nurses (both in the classrooms and at the clinical skill labs), videoconferencing devices and the communication medium that supplies the interaction between students and the nurses. The image quality must be clear enough to allow students’ detailed views of the procedures, for instance, to distinguish between an organ and a vein. Hence, the video system should be High Definition (HD). At the skill lab, a high resolution (HD) video camera and a sound system are installed, and HD video frames with nurse’s voice are sent to classrooms or individuals via LAN over the Internet. Here, the quality and the speed of the communication are very important, since approximate size of zipped 1 second of the HD video frames is 1 MB; thus the connection speed (if digital communication is used) should be at least 10Mb/sec. The classrooms are equipped with a projection system to show the video and sound data coming from skill lab, and a wireless microphone to enable students to ask questions. In fact, a HD video camera system may be installed in the classrooms, which enables the nurse in the skill lab to monitor the classrooms and students. However, we do not recommend bidirectional video interaction, because this may distract the nurse from the procedure.

Experiment results and data analysis

Experimental setup

In this study, an experiment was conducted to teach ECG device features. Our choice of electrocardiography is mainly due its visual nature and the need to resolve the confusion that commonly occurs while placing the electrodes. We have implemented a very low cost setup for our study using, private lines to transmit the data. Fig.1 shows the setup used in this experiment. The primary reason for using this was insufficient transmission bandwidth allocation at the location. An IP based solution would equally work well. As shown in Fig. 2, a clinical skill lab was designed with the following characteristics: a HD video camera, a sound system, an ECG device, a student volunteer pretending to be a patient and a nurse (Registered Nurse, with PhD). The wired connection was used as a communication medium to increase the video data transmission quality. At the classroom end, we used a projection system, a notebook computer for lecture slides, a lecturer (RN, PhD), and a wireless microphone to be used by students for their questions (to interact with nurse in the clinical skill lab). To avoid potential anxiety, we choose not to put the students themselves “on camera.”

![Figure 1. Videoconferencing setup for the experimental study](image-url)
Data collection

The research was conducted at EGE University which is one of the largest universities in Turkey with the participation of 13 male and 57 female second year students studying at the nursing department. The nursing department is a four-year program that offers a Bachelor of Science degree and has approximately 400 students. The university accepts its student based on a nationwide university entrance exam, and the nursing department accepts its students from the 5% percentile based on the past 10 years data. The research is conducted as part of a lecture which is compulsory for sophomores.

To ensure the homogeneity of both control and experimental groups, an even distribution of cumulative point averages and gender were taken into consideration when assigning students to groups. Almost all students were aware of current technologies available for interactive education and were current users of those technologies (Table 1). Moreover, the ECG device and its features were selected for teaching. In the experiment, a test that was administrated at three different times, as pretest, posttest I and posttest II, to assess the effect of the proposed model, and a questionnaire was also given to obtain students’ opinion about the videoconference-supported lecture.

<table>
<thead>
<tr>
<th>Table 1. Computer and information use of students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Owns a computer</td>
</tr>
<tr>
<td>Access to internet</td>
</tr>
<tr>
<td>Used internet for literature survey</td>
</tr>
<tr>
<td>Attended videoconferencing lectures in university education</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Procedure

The experimental and the control groups were given a pretest assessment that consists of a questionnaire on socio-demographic data, including age, educational status, sex and prior computer and informatics knowledge, followed by a twenty multiple choice question knowledge test to determine ECG knowledge level. Each student test is graded according to a 0-100 scale. After the completion of the test, both groups were given a 40 minute power point presentation on the electro physiology of the heart, and the classification and types of arrhythmias. The lecture included information regarding the ECG device and description of the procedure. Following the lecture, the experimental group established a video link to the skill lab for the introduction of the ECG device by a nurse who described and explained the features of the device and the probes connected to the device and their color coding.
Next, students watched a simulation of an ECG procedure performed on the patient. The nurse answered student questions during and after the procedure and the session was closed. The control group was re-given the knowledge test again immediately after the lecture, and the experimental group was given the same test following the videoconferencing session. The students were also given another post-test questionnaire to find out student opinion about videoconference-supported lecture. The posttest knowledge test was administrated to both groups two weeks later.

Data were analyzed using the statistical software Statistical Program for the Social Sciences SPSS for Windows version 15.0. Baseline characteristics of control (TCG) and experimental (VCG) were compared by Student’s t-test. ECG knowledge level baseline and post-rehabilitation data were compared in each group with repeated-measures analysis of variance and covariance

**Results and analysis**

Table 2 shows there were no significant differences in mean ECG knowledge scores between the control (TCG) and experimental (VCG) groups at baseline (t = 0.187, p = 0.852). Although the students had received no previous official training for the ECG device before the experiment, the pretest scores for both groups were reasonable because of their personal interest in ECG.

**Table 2. Student t test for baseline characteristics of VCG and TCG**

<table>
<thead>
<tr>
<th>ECG Knowledge Mean</th>
<th>TCG (Traditional Classroom)</th>
<th>VCG (Video-Conference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>X</td>
<td>SD</td>
</tr>
<tr>
<td>--------------------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>ECG Knowledge Mean</td>
<td>34</td>
<td>46.17</td>
</tr>
</tbody>
</table>

We applied Shapiro-Wilk normality test to the results. Discrepancies in normality were found only in topic 2 and 3 (VCG and TCG percentages), and no meaningful normality discrepancy was detected in the others. Furthermore, we tested results with Friedman-Wilcoxon and Mann Whitney U., and the findings indicated that parametric methods are more suitable. Besides, since showing changes between groups’ test results in the time domain (interaction) is important, we have chosen Repeated Measure ANOVA test for statistical analysis.

The results of Repeated Measure ANOVA analysis showed that there were significant improvements in ECG knowledge scores in the VCG group (f = 5.67, p = 0.004) (Table 3).

**Table 3. Change in knowledge level, in the Videoconference (VCG) and Traditional Education Group (TCG) groups**

<table>
<thead>
<tr>
<th>GROUP</th>
<th>TIME</th>
<th>N</th>
<th>X</th>
<th>SOURCE</th>
<th>Mean Square</th>
<th>df</th>
<th>f</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCG</td>
<td>Pre test</td>
<td>34</td>
<td>46.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post test I</td>
<td>34</td>
<td>76.76</td>
<td></td>
<td>24659.52</td>
<td>2</td>
<td>144.37</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Post test II</td>
<td>34</td>
<td>64.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VCG</td>
<td>Pre test</td>
<td>36</td>
<td>45.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post test I</td>
<td>36</td>
<td>88.61</td>
<td>TIME*</td>
<td>969.52</td>
<td>2</td>
<td>5.67</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Post test II</td>
<td>36</td>
<td>76.52</td>
<td>GROUP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>170.80</td>
<td>136</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Covariance analysis results were conducted to understand whether these differences occurred across time periods, and results showed significant differences between groups for both posttest I (f = 15.08, p < 0.05) and posttest II (f = 8.38, p < 0.05) in terms of test scores (Table 4 and 5).

**Table 4. Change of knowledge level between pretest and posttest I, in the Videoconference (VCG) and Traditional Education Group (TCG) groups**

<table>
<thead>
<tr>
<th>GROUP</th>
<th>TIME</th>
<th>N</th>
<th>X</th>
<th>SOURCE</th>
<th>Mean Square</th>
<th>df</th>
<th>f</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCG</td>
<td>Pre test</td>
<td>34</td>
<td>46.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post test I</td>
<td>34</td>
<td>76.76</td>
<td></td>
<td>2543.83</td>
<td>1</td>
<td>15.08</td>
<td>0.000</td>
</tr>
<tr>
<td>VCG</td>
<td>Pre test</td>
<td>36</td>
<td>45.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post test I</td>
<td>36</td>
<td>88.61</td>
<td>TIME* GROUP</td>
<td>2543.83</td>
<td>1</td>
<td>15.08</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>168.68</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Change of knowledge level between pretest and posttest II, in the Videoconference (VCG) and Traditional Education Group (TCG) groups

<table>
<thead>
<tr>
<th>GROUP</th>
<th>TIME</th>
<th>N</th>
<th>X</th>
<th>SOURCE</th>
<th>Mean Square</th>
<th>df</th>
<th>f</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCG</td>
<td>Pre test</td>
<td>34</td>
<td>46.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post test</td>
<td>34</td>
<td>64.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VCG</td>
<td>Pre test</td>
<td>36</td>
<td>45.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post test</td>
<td>36</td>
<td>76.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TIME*** GROUP 333.44  1  8.38  0.005

Error 334.44  67  -  -

The student satisfaction with the videoconferencing was high as per the results of a questionnaire given after posttest II. Table 6 shows the related questions and the students’ responses. The most significant results were: All students (100%) reported finding the videoconference-supported lecture an enjoyable activity, and all (100%) reported that they would recommend others to attend this type of lecture. The results also indicated that almost no distraction or loss of concentration (97.2%) during the lecture, and that for most (80.6%), it did not hinder their ability to ask questions. None thought the videoconference session was excessively long. Finally, the students’ overall satisfaction level showed that 94.4% were highly satisfied with videoconference-supported lecture.

Table 6. Questionnaire to determine student opinion about videoconferencing

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
<th>Partially</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- I was satisfied with the videoconference-supported lecture.</td>
<td>33 (91.7%)</td>
<td>-</td>
</tr>
<tr>
<td>2- I want to have more videoconference supported lectures.</td>
<td>33(91.7%)</td>
<td>-</td>
</tr>
<tr>
<td>3- Sound quality was a problem.</td>
<td>2(5.6%)</td>
<td>25(69.4%)</td>
</tr>
<tr>
<td>4- Video quality was a problem.</td>
<td>-</td>
<td>29(80.6%)</td>
</tr>
<tr>
<td>5- The classroom environment was conducive to learning overall.</td>
<td>27(75%)</td>
<td>2(5.6%)</td>
</tr>
<tr>
<td>6- Videoconference-supported lectures positively contributed to my education.</td>
<td>34(94.4%)</td>
<td>-</td>
</tr>
<tr>
<td>7- Videoconferencing hindered my ability to ask questions.</td>
<td>1(2.8%)</td>
<td>29(80.6%)</td>
</tr>
<tr>
<td>8- Having more than one instructor adversely affected my concentration.</td>
<td>-</td>
<td>35(97.2%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At the beginning</th>
<th>At the end</th>
<th>No preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>9- Which part of the lecture that videoconferencing should be carried out?</td>
<td>11 (30.6)</td>
<td>15(41.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Too Long</th>
<th>Too Short</th>
<th>Adequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>10- What do you think about the duration of the videoconference session?</td>
<td>-</td>
<td>11(30.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
<th>No opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>11- I have attended videoconference-supported lectures before.</td>
<td>5(13.9)</td>
<td>29(80.6)</td>
</tr>
<tr>
<td>12- I would like to watch recorded lectures if they are available.</td>
<td>35(97.2)</td>
<td>1(2.8)</td>
</tr>
<tr>
<td>13- I recommend videoconference-supported lectures to others.</td>
<td>36(100)</td>
<td>-</td>
</tr>
<tr>
<td>14- I lost concentration because of videoconference.</td>
<td>1(2.8)</td>
<td>35(97.2)</td>
</tr>
<tr>
<td>15- Videoconference-supported lectures are significantly better than traditional ones.</td>
<td>35(97.2)</td>
<td>1(2.8)</td>
</tr>
<tr>
<td>16- Knowledge retention is higher in videoconferencing compared to traditional education.</td>
<td>35(97.2)</td>
<td>1(2.8)</td>
</tr>
<tr>
<td>17- Videoconference is a distracting activity.</td>
<td>2(5.6)</td>
<td>34(94.4)</td>
</tr>
<tr>
<td>18- Videoconference makes lectures enjoyable.</td>
<td>36(100)</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Not Satisfied</th>
<th>Neither satisfied nor dissatisfied</th>
<th>Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>19- Your overall assessment of videoconference supported.</td>
<td>-</td>
<td>2 (5.6%)</td>
</tr>
</tbody>
</table>
Discussions and conclusion

In this study, we conducted an experiment to investigate the effects of using videoconferencing as a part of e-learning activity on the ECG skills of the 2nd year nursing students. The results showed that the videoconference assisted group demonstrated a higher level of ECG knowledge than the traditional lecture based class. This result is consistent with the findings of Jang et al. (2005), who used web-based instructions to teach ECG to nursing students. However, our results contrast with those of Buckley, who found no difference in course grades or test scores from those of traditional classroom courses. Unlike the present study, Buckley had no aspect of real time nurse-student interaction. This suggests that visual stimulation and interaction is a key component in learning, and may play an effective role in nursing education in terms of knowledge acquisition. This result is in accordance with the findings of Zerr et al. (2008). On the other hand, the posttest II, administrated with the purpose of measuring the ECG knowledge retention rate, showed that the rate of decrease in scores were approximately same for both groups; thus we can say that there is no significant difference between the control and experimental groups in terms of knowledge retention, and that the effect of videoconference-supported lectures on knowledge retention is negligible. This outcome agrees with that of Buckley (2003), Jeffries et al., (2003), and Leasure et al., (2000), who all used web based courses in their studies.

We administrated two different questionnaires. The first one was given to both groups before the experiment to collect demographical profiles. The other questionnaire was given only to the experimental group after the experiment, to assess the effects of the videoconference-supported lecture. Results showed that the almost all students were aware of current technologies available for interactive education and were current users of those technologies (Table 1). Therefore, they were open and receptive to using interactive learning technologies. Furthermore, the final questionnaire results showed that the overwhelming majority in the experimental group were content to participate in the experiment and were satisfied with the videoconference-supported lecture. Our observations indicated that the students found the activity enjoyable, and were able to follow the lecture throughout the experiment. This can be considered a beneficial outcome, in view of the fact that the students tend to lose concentration during more traditional lectures. The student opinion about videoconferencing assisted course supports the results of a similar study by Rush et al. (2011), who used videoconferencing in a diabetic clinic.

Several other indicators of student satisfaction were observed during the experiment. Students in control group asked fewer questions during the lecture, while in the experimental group, many perceptive questions were asked; including what should be done in the case of a patient with an amputated extremity, or injury. We believe that such questions demonstrate the positive effects of visual stimuli in lectures. One possible benefit of the system is that interactive learning enables the nursing students to observe the clinical environment and nurse-patient interaction before their actual clinical practice, which can increase the effectiveness of communication with the patient, moral knowledge development, ethical development and caring attributes. This is especially important where every patient has different conditions and attitude toward medical care. Students can theorize about various conditions about the patient, and the nurse can demonstrate through simulation how to handle such conditions. Placing the ECG probes in a different part of the body because of an injury is a good example of handling an uncommon situation. However, the ECG process used in the study is designed under the assumption of the absence of the secondary disorders (non-comorbidity).

There were a number of lessons learned during the experimental study period. Most importantly, the nurse at the skills lab should not become distracted from the task because of this teaching activity, therefore one directional video transmission is recommended to minimize possible negative effects on the nurse, such as anxiety and loss of focus.

The experiment conducted in this study revealed that interactive learning methods have several restrictions and disadvantages. Technology and cost related issues are important factors needing effective management for successful integration of technology in the classrooms. A study by Potempa et al. (2001) indicated the leading deterrents were cost and technology skills training, factors also valid for our study. Although internet connection is widely available, it may not always have the sufficient bandwidth to allow the quality of video data transmission required. In this case, the video frame rate or the size can be reduced to catch seamless transmission, or other communication alternatives can be considered, such as connecting the skills lab with an Audio-Video cable or setting up a local wireless connection using WIFI or a wired connection using Local Area Network (Ethernet). These methods are simple to implement and provide immunity from internet related connection and bandwidth problems. Another concern is the cost of technical infrastructure. The quality of the communication and the cost are proportional, and higher quality
means higher cost. On the other hand, it is worthwhile to note the fact that cost of technology is inversely proportional to time passed, since the adaption of the technology. Moreover when it comes to health, an increase in educational costs is tolerable when the return is higher. One other issue with the videoconference supported lecturing is that finding an available clinic and a nurse at the time of lecture is not always possible; thus a dedicated skill lab and an extra lecturer/nurse maybe required. Since the nursing departments usually have skill labs which is also the case in our study, this issue simply reduces to scheduling and allocation of the skill lab. As far as the need for an extra lecturer is concerned, again the problem can be addressed by reallocating the time of the two lecturers. Instead of two lecturers each teaching their theoretical and practical classes separately, they could divide the work up. In a more cost effective approach multiple instructors work with the same clinical nurse. In traditional classroom teaching, the lecturer simply prepares the materials regarding the course subject; however in interactive videoconference-supported lectures, both lecturer and nurse need to cooperate on the production of course materials. Thus this may mean allocation of extra time.

References


Predicting Adolescent Deviant Behaviors through Data Mining Techniques

Yu-Chin Liu and Yung-Chieh Hsu
Department of Information Management, Shih Hsin University, Taipei, Taiwan 116, R.O.C. // ycliu@cc.shu.edu.tw // alain.hsu@msa.hinet.net

*Corresponding author

(Submitted August 17, 2011; Revised February 1, 2012; Accepted March 13, 2012)

ABSTRACT
Adolescence is the time during which people develop and form their crucial values, personality traits, and beliefs. Hence, as deviant behaviors occur during adolescence, it is important to guide adolescents away from such behaviors and back to normal behaviors. Moreover, although there are various kinds of deviant behavior, most of them would either directly or indirectly affect youths’ academic learning progresses. Therefore, many researchers have endeavored to explore the issues of juvenile delinquency. In this study, we focus on providing methods that could assist counseling officers in discovering symptoms and pre-symptoms of youth delinquency. The study proposes a framework for mining associations from “memo-type” records without supervision of senior experts. In addition, we apply this approach to retrieve associations among deviant behaviors from the counseling records stored in databases. The contributions to adolescent counseling are as follows: (1) A keyword tree of deviant behaviors is constructed and verified; (2) the general and time sequence associations of deviant behaviors are mined from memo-type records and also verified. Finally, (3) an information system is recommended to help junior counselors performing counseling works. Consequently, without experienced consultants, the proposed framework could discover valuable knowledge from consulting records effectively and many efforts are therefore saved.

Keywords
Data mining, Association analysis, Adolescent deviant behavior, Counseling record

Introduction
In past, research (Freisthler, Byrnes, & Gruenewald, 2009; Soenens, Vansteenkiste, Smits, Lowet, & Goossens, 2007) has pointed out that if deviant behaviors cannot be indentified and corrected as they occur, adolescents might do something uncontrollable or unredeemable later in their lifetimes. Therefore, in order to avoid consuming social resources in future, deviant behaviors had better be detected and corrected as early as possible. For this reason, many researchers strive to study the prevention and amendment of adolescent deviant behaviors to avoid further serious social problems. Further, some researchers also discuss how to help the young devote themselves to meaningful activities and goals.

From the perspective of educators, the deviant behaviors might greatly influence students’ learning progresses since some deviant behaviors are directly correlated in learning, such as: negative learning attitude, truancy, and disobedience. Others might result in addictions or bad relationships; therefore students could hardly focus on academic learning. Therefore, preventing adolescents from deviant behaviors could also improve their learning progresses and abilities to acquire knowledge.

As information technology prevails worldwide, electronic databases have been used to store mountains of data with regards to various fields. Information systems are efficient at collecting data as well as transforming data into useful information. There are many IT tools which can integrate data from different sources, so that the cost of information processing can be reduced and resource allocation can be much more efficient. So with regards to supervising students’ deviant behaviors, archived counseling records are very helpful for counselors to trace, monitor, and record behaviors, treatments, and outcomes.

In general, the records stored in database can be in the form of structured or unstructured data fields. The structured fields refer to data which can be briefly summarized as values and stored in certain fields. Contrarily, unstructured fields hold data which is difficult to categorize as finite sets of values and therefore stored intact. Comparing to structured fields, unstructured data fields are difficult to understand and be utilized.
In this paper, we aim to propose a framework to analyze the unstructured but crucial text in databases. We term such unstructured data as “memo-type” data throughout this paper. Such memo-type data is usually presented in the form of short paragraphs, but these paragraphs are dissimilar to regular articles. “Memo-type” records exist in many real applications, such as patients’ diagnosis and treatment data, the counseling records kept by psychiatrists, the maintenance records for hardware equipment, and answers from open questions, and working diaries or journals. While it is known that the text stored in memo-type fields is good at providing details, data utilization is regrettably hampered.

Traditionally, in order to retrieve information from memo-type fields, experts need to transform them into structured data according to domain knowledge. However in practice, such methods encounter certain problems. Firstly, the transformation relies heavily on expert intervention, but various experts have different opinions. Secondly, the newly transformed fields designed by domain experts often result in data records that are too sparse to maintain. Hence, to develop data records which are simple in nature but within which major information is stored in memo-type fields, the method stated above is unsuitable.

In summary, due to the characteristics of memo-type records, traditional document mining techniques are untenable, and the mining task for these memo-type fields should be handled with a more focused search strategy. To the best of our knowledge, no literature focuses on mining associations from “memo-type” records; therefore, in this research, we propose a semi-automatic approach to analyze memo-type records with only little expert interventions. By applying our method, research results can help practitioners classify records and make use of information more easily.

To illustrate our proposed method, we analyze the electronic counseling records stored in databases. By applying this method, junior counselors can mine associations among deviant behaviors from counseling databases and can then further predict possible deviant behaviors which might happen in the future or already have occurred but unnoticed yet. Moreover, young school officers can therefore use these findings to guide students back to normal behaviors timely without the help of senior counselors. Namely, with such knowledge, counseling work could be done much more effectively.

The two major types of associations considered in our study are explained as followed. The first one is to mine for the concurrence of deviant behaviors. This information is very useful in understanding which deviant behaviors are likely to occur at the same time and to further alert the counselors if any other possible deviant behaviors are concealed during consultations. Consequently, junior counselors can provide sufficient support to teenagers timely, and deviant consequences can be further avoided.

The second technique is to discover the time-ordered associations of deviant behaviors. The time sequence associations represent the findings in the form of “If some deviant behaviors occur, and then certain deviant behaviors will happen next.” Once such sequence associations are discovered, they can be used to predict possible deviant behaviors that might happen in near future. Furthermore, they are very helpful in steering young offenders away from committing other delinquent acts.

To achieve the goals stated above, our proposed approach is first to construct a keyword tree containing terms related to adolescent deviant behavior. The keyword tree is then used to classify computerized counseling records automatically. The detail steps of constructing the keyword tree will be explained. Traditionally, such keyword tree is built by domain experts. They first discern and define various kinds of deviant behaviors, thereafter individual records are classified into proper categories manually. Although such approach can achieve a high degree of accuracy, it is difficult to apply to large volume of data. In addition, since the work is carried out by domain experts, the classifications made could be subject to their various viewpoints. Therefore on contrarily, our proposed semi-automatic method could categorize data records much more objectively.

The rest of the paper is organized as follows. Literatures are reviewed in the second section. The third section details the proposed framework as well as the procedures for analyzing memo-type data. The fourth section first verifies the effectiveness of the keyword tree constructed through this proposed method. Further, the associations among adolescent deviant behaviors are discovered and also validated. Finally, conclusions and suggestions are drawn in the fifth section.
Related Work

In order to mine associations among deviant behaviors, first, literatures on youth deviant behaviors and psychological counseling systems are reviewed first. Second, data mining techniques including text mining, association discovery and time sequence analysis are surveyed.

Youth deviant behaviors and psychological counseling systems

The definition of a “deviant” refers to one who differs from a norm, especially a person whose behavior and attitudes is considered to be unacceptable (Lin, 1991). Deviant behavior includes eccentric, abnormal, criminal, maladaptive, and anti-social behaviors. More specifically, adolescent deviant behavior refers to behavior which a normal youth would not exhibit. It comprises both problems such as learning disabilities, and the maladaptive behavior caused by physical disabilities or emotional/behavioral disorders.

Some primary and secondary counseling related research has classified youth delinquency into bad habits and extroversion delinquency (Liao, 2002). Drug abuse, sexual problems, and harmful amusement have been classified as bad habits, while truancy, breaking school rules, resisting authority, running away from home, assault and battery, and theft have been classified as extroversion delinquency. Since such classifications coincide with the goals of our research, we have defined our keyword tree by referencing the deviant behaviors defined in related research. These behaviors are summarized in Table 1.

<table>
<thead>
<tr>
<th>Classifications</th>
<th>Deviant behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outward behaviors (extroversion/externalizing behaviors)</td>
<td>truancy, running away from home, noncooperation, resistance, disobedience, untruthfulness, ruin, ravage, menace, bullying.</td>
</tr>
<tr>
<td>Inward behaviors (introversion/internalizing behaviors)</td>
<td>recoil, negative noncooperation, overdependence, day-dreaming, anxiety reaction, masochism, self-mutilation.</td>
</tr>
<tr>
<td>Academic problems</td>
<td>unwillingness to do homework, cheating in tests, opportunism, free-riding, inadvertent, attention-deficit, unsteady academic progress.</td>
</tr>
<tr>
<td>Eccentric habits</td>
<td>sucking thumbs, biting fingernails, twitches, stammering, ill-balanced feeding, bed-wetting, substance abuse, sexual deviance.</td>
</tr>
<tr>
<td>Anxiety disorders</td>
<td>neurotic disorder/behavior, such as tenseness, shivering, feeling nausea, heart/chest illness, general weakness, anxiety, obsessive compulsive disorder, hysteria.</td>
</tr>
<tr>
<td>Psychosis</td>
<td>psychopathology, schizophrenia, manic-depression.</td>
</tr>
<tr>
<td>Drug abuse</td>
<td>smoking, drinking, betel nut chewing, doping, nighthawk, snorting LSD, pent azocine injection, sniffing glue, etc.</td>
</tr>
<tr>
<td>Harmful amusement</td>
<td>billiards and pool, gambling, video-game addiction, playing Mahjong, lingering in dance halls or adult cafes.</td>
</tr>
<tr>
<td>Problem behaviors concerning sex</td>
<td>biased sex ideas, opposite sex infatuation, escort, dirty jokes, pornographic books/periodicals/videos, sexual abuse, fondle bodies.</td>
</tr>
<tr>
<td>Transgression of school rules</td>
<td>talkativeness, creating distractions, poor peer relationships, aversion to school, lateness, leaving early, absence from class, truancy.</td>
</tr>
<tr>
<td>Resisting authority</td>
<td>rebellious attitudes, turning in homework late, cheating, deceptive behavior, talking back.</td>
</tr>
<tr>
<td>Running away from home</td>
<td>wandering outside after school, intention to run away, running away from home.</td>
</tr>
<tr>
<td>Assault and other delinquency</td>
<td>fighting, bullying, impoliteness, untruthfulness, taunting, damaging public property, ganging, threatening.</td>
</tr>
<tr>
<td>Theft and plunder</td>
<td>keeping found money, not returning borrowed money, stealing money, looting wealth and properties, extorting money.</td>
</tr>
</tbody>
</table>

To guide students away from delinquent behavior is an indispensable goal of school education; therefore school officers strive to build effective guidance systems. In addition, it is well-known that if families cooperate closely
with schools, adolescent deviance in schools can be greatly reduced. Traditionally, school officers have spent most of their time on individual client evaluation. Recently, however, most studies of guidance practices claim that the school authorities in order to save time and money should focus on prevention and early intervention before deviant problems become more complicated and insolvable. After all, this is the best way to save resources and achieve beneficial results.

Information technology has shown its ability to allow for knowledge discovery and information retrieval across many applications, and school officers can integrate IT tools into guidance work for greater utilization of resources. In actuality, IT technology could help student guidance and counseling work in many ways, including through communication, information sharing/dissemination, and the discovery of knowledge. Great achievements in the field of e-learning have already been made, but there is still room to further make use of IT technology in the field of student guidance (Tsau, 2008). For example, IT tools could be used to analyze electronic counseling records to determine common behaviors from mountains of data. Furthermore, the knowledge discovered could contribute to the work of prevention, consultation, and intervention. IT technology could be also very helpful in supporting the practices of school guidance work, and presents itself as a fascinating topic for research.

**Knowledge discovery**

Knowledge Discovery is a process to extract valid, novel, and useful information from large amounts of data. Major techniques in data mining include basket analysis (Agrawal & Srikant, 1994), data classification (Quinlan, 1993), and data clustering (Wang, Xu & Liu, 1999). In addition, the processes of analyzing text data are known as “Knowledge Discovery from Text” or “Text Mining,” which generally refers to retrieving valuable knowledge from unstructured text. Related literatures are as followed.

**Association analysis**

Association analysis has received great attention in recent years. Due to its ease of comprehensibility and application, it has been frequently used to analyze data in all kinds of applications and is recognized as one of the most important data mining techniques. It was first introduced to retrieve the associations among items sold and was used to acquire strong association rules shown in the form of “If one buys x, then he/she buys y.” (Agrawal, Imielinski & Swami, 1993) Such applications are very helpful to store managers for positioning the right products for the right customers.

Thereafter, various kinds of association applications; and among them, sequential mining (Srikant & Agrawal, 1995) aims to discover events happened in time sequence. A sequence rule is shown as “If one buys x, then he/she will buy y next.” Such that the products which customers willing to buy next could be predicted.

**Text mining**

Text Mining is defined as editing, organizing and analyzing amounts of documents to discover specific features and associations (Sullivan, 2001). It refers to handling text data through several techniques such as information retrieval/searching, data mining, machine learning, statistics, and natural language. In contrast to other techniques; Text Mining requires additional steps to pre-process text documents. In general, Text Mining includes several techniques, such as feature extraction, keyword retrieval, context searching, document auto-categorization/clustering, and auto-summary. In the following paragraphs, we briefly introduce the concepts of document clustering and feature extraction since they are relevant to our research.

Document Clustering in Text Mining is used to segment documents into several clusters, and then people could label each cluster with a proper subject. Although document clustering techniques may not achieve complete accuracy, they help people to process large amounts of documents automatically and efficiently. In order to cluster documents more reliable, the feature extraction techniques should first identify important words from documents as well as to identify key words/phrases with the same meanings but different forms. Then, the distinguishable features could be retrieved as clustering attributes.
Commonly, the article in a document is longer and more detailed than the text stored in memo-type fields. Unlike regular articles, which are the focus of these text mining techniques, memo-type data is presented in the form of short paragraphs and contains more keywords and commonalities. In contrast to regular documents, the content of memo-type fields is usually much more focused on specific events and contains less variety. Therefore, traditional text mining techniques could be further refined to mine valuable knowledge from memo-type data.

Research methodology

In this study, we propose a framework to mine associations from memo-type records. The methodology can be divided into two phases. In Phase I, the keyword retrieval techniques would be used to discover key words which represent the main characteristics of memo-type data. In Phase II, the counseling records are replaced with these keywords first and later the associations and time sequence associations among keywords are discovered. For our sample, since the co-occurred deviant behaviors are discovered, junior counselors could therefore help teenagers away from them timely.

In this section, the example of deviant behaviors stored in student counseling records is used to illustrate our proposed method. The research framework is first shown in Figure 1. Second, the background of the example is introduced first, and later definitions of the keyword tree are explained. Last, the details of Phase I and II in Figure 1 are illustrated.

*Figure 1. The research framework for mining associations among deviant behaviors*
The source database and related documents

In this study, the data source comes from an information system which records counseling events occurred at senior high schools in Taiwan, R.O.C. All data is collected by Military-trained Nursing Instructors and is kept by the Education Administration of Taipei Government. Although there might be other school-based information; since it was not accessible in this study due to issues of privacy, we therefore focused on the collected records.

A total of 60,907 records had been collected from January, 2007 to October, 2008, and they had originally been classified into 10 categories as shown in Table 2. Of the 60,907 records, only 32,908 of them are relevant to deviant behavior. Hence, only these 32,908 counseling records would be analyzed.

<table>
<thead>
<tr>
<th>Category</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referral Service</td>
<td>230</td>
</tr>
<tr>
<td>Contact with Parents</td>
<td>5,781</td>
</tr>
<tr>
<td>Visit Accommodations Outside Campus</td>
<td>341</td>
</tr>
<tr>
<td>Accident Event Handling</td>
<td>220</td>
</tr>
<tr>
<td>Services for Students of Other Schools</td>
<td>107</td>
</tr>
<tr>
<td>Emergency Rescue Services</td>
<td>416</td>
</tr>
<tr>
<td>Health Care</td>
<td>1,880</td>
</tr>
<tr>
<td>Emotion Control Management</td>
<td>4,474</td>
</tr>
<tr>
<td>Special Events Treatment</td>
<td>4,040</td>
</tr>
<tr>
<td>Others*</td>
<td>43,418</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>60,907</strong></td>
</tr>
</tbody>
</table>

* A record could be classified into more than one category.

As shown in Table 2, although each record has already been categorized, there are still 23,306 of the records classified as “Others.” Namely, 70.08% of the counseling records cannot be classified properly, and then these records would be very difficult to analyze and use. In order to ensure that counseling records are more comprehensive and applicable, new classification taxonomy should be developed.

Further, in general while performing counseling, there should be official guidelines for counselors to follow and such documents are referred as relevant documents in this paper. Therefore, the present study, we adopt the documents stating the definitions of deviant behavior from the list of Classified Campus Safety Events, which are issued by the Education Department of Taipei Government.

During the Data Pre-processing and Cleaning step, the source database and the related documents described above were first reviewed, cleaned, and retrieved. Therefore, the keywords could be recognized and used to construct the keyword tree. In this database, the selected data fields include Date, Time, StudentID, Counseling Category, FIRST and Content. Among them, the Content field details the counseling events briefly and is a typical memo-type field. Some data examples from the database are shown in Table 3.

Table 3. The counts for each category of the collected data records.

Table 3. Examples of the counseling records stored in the database.

<table>
<thead>
<tr>
<th>Counselor</th>
<th>Consulting Date</th>
<th>Student Name</th>
<th>Category</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>2009/3/3</td>
<td>-</td>
<td>Others</td>
<td>該生上課時違反教室規定—聽隨身聽，輔導其上課應該專心聽講。This student listened to music on an iPod in classes, so we guided him not to do so again.</td>
</tr>
<tr>
<td>-</td>
<td>2009/3/2</td>
<td>-</td>
<td>Others</td>
<td>該生近日常缺曠，輔導其應專心於課業，不要將心力用於其它事物上，以致缺曠不來上學。This student was absent from classes these days. We told him of the importance of attending classes regularly.</td>
</tr>
</tbody>
</table>

*FIRST field is used to indicate whether the client being recorded in the database for the first time.
Although there are many procedures in the proposed framework, they mainly consist of two phases: Keywords Analysis and Association Rule Mining, which are later explained.

**The keyword tree**

To demo the proposed method, the keyword tree is first explained. Although later we demo the proposed method by analyzing the consulting records according to the constructed 3-level keyword tree, it is certainly not limited to three levels. Hence, we generalize the keyword tree as $k$ levels ($k \geq 2$), it is defined as a tree composed of two parts: the top ($k-1$) levels which present the domain-specific categorical hierarchy built by domain-experts as well as the $k^{th}$ level list which contains keywords (filtered Chinese terms) related to their corresponding ($k-1)^{th}$ sub-categories. Figure 2 illustrates the keyword tree with $k = 3$. The top 2 ($k-1$) level of the keyword tree in Figure 2 refers to the classification taxonomy which is domain-dependent; so that it might be very common in its related domain and can be recognized by experienced domain-experts without difficulties. Furthermore, the $k^{th}$ level of the keyword tree is very essential since our proposed framework uses it to categorize data records as the sub-category shown in its parent ($k-1$) level. Briefly, the keywords (filtered Chinese terms) in $k^{th}$ level bridge memo-type records into real and meaningful ($k-1)^{th}$ level sub-categories for practitioners. For example in Figure 2, if a memo-type record contains the keyword $KW_{211}$, then it would be classified as sub-category $C_{21}$.

![Figure 2. A sample of 3-level keyword tree](image)

**Phase I: Keyword analysis**

In this subsection, steps for Keyword Analysis are introduced. The main goal of this phase is to analyze memo-type data and indentify meaningful keywords. Initially, the original data is cleaned and fed into the proposed framework. Phase I would therefore produce the keyword tree to be used for associations mining in Phase II. Simultaneously, in order to produce a better keyword tree, the proposed framework also recommends applying the same analysis procedures to relevant documents which state domain-specific regulations and principals. Since the keywords identified from relevant documents are significantly focused on our specific application, they are good candidates to be included in the keyword tree.

Further, since our data is in Chinese, the problem occurred first is to parse Chinese words from paragraphs. Unlike English, in which every word has a clear definition, a meaningful Chinese word usually composes of more than one word to form certain meaning, and there is no blank space between words. Therefore, in this step, a Chinese parser system needs to be used to analyze and segment Chinese words in data records.

The parser system adopted in this study is developed and maintained by the Chinese Knowledge Information Group (CKIP, 2011). This well-known Chinese text parser has been widely used in mining Chinese text (Lu, Chien & Lee, 2002, 2004; Yang & Lee, 2004). Since the Term Frequency and Inverse Document Frequency (TF-IDF) is a widely adopted measure to screen keywords from documents (Salton & Buckley, 1988), therefore words identified by CKIP parser are further examined by calculating TF-IDs and also by consulting domain experts. Finally, the keywords are filtered and the keyword tree is finally constructed.
In our running case, the Chinese parser returns a list of Chinese words by analyzing the database and relevant documents, respectively. Then, TF-IDF values are applied to screen crucial keywords from all parsed Chinese words. According to the qualified keywords, the keyword tree is therefore constructed. Finally, since the mining results of our proposed framework highly rely on the keyword tree, it is necessary to validate the constructed keywords tree before performing further mining tasks. The constructed tree in our running case contains three levels: the 1st level consisting of 10 categories and the 2nd level consisting of 43 sub-categories. Table 4 lists the category in the 1st level of constructed keyword tree.

Table 4. The total number of records categories as 1st level of the keyword tree.

<table>
<thead>
<tr>
<th>First Level Keywords</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Family issues (parental marriage status)</td>
<td>275</td>
</tr>
<tr>
<td>B. Extroverted behaviors (additions, transgression of school rules, committing crimes)</td>
<td>14,917</td>
</tr>
<tr>
<td>C. Deviant behavior in learning (negative learning attitude, truancy, disobedience)</td>
<td>11,528</td>
</tr>
<tr>
<td>D. Deviant behavior after school (car racing, harmful amusement, running away)</td>
<td>4,421</td>
</tr>
<tr>
<td>E. Sexual deviance problems</td>
<td>315</td>
</tr>
<tr>
<td>F. Introverted behaviors (emoted barriers)</td>
<td>4,729</td>
</tr>
<tr>
<td>H. Disciplinary</td>
<td>4,056</td>
</tr>
<tr>
<td>I. Positive incentive (administrative rewards)</td>
<td>815</td>
</tr>
<tr>
<td>J. Follow-up tracking (probations, transfer)</td>
<td>2,775</td>
</tr>
<tr>
<td>R. Re-entry</td>
<td>179</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>44,010</strong></td>
</tr>
</tbody>
</table>

Phase II: Association rule mining

Based on the keyword tree, each record is first summarized by transforming the contents into the keywords shown in the kth level of the tree. Table 5 presents one of the transformed data records. The keyword’s code is meaningful, the leftmost letter representing the 1st level category of the keyword tree and the number referring to the corresponding 2nd level sub-category. For our study, the numbers of records classified as the 1st level of the keyword tree are presented in Table 4.

Table 5. A sample of transformed data record.

<table>
<thead>
<tr>
<th>ID</th>
<th>Content</th>
<th>Keyword Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-217</td>
<td>該員於校外網咖內抽箊為警方查獲，相關公文到校後，找該員至教官室瞭解相關情形，然後告知該員抽箊的害處與校規相關之規定，並告知導師與通知家長。</td>
<td>D1 B3 (網咖，抽箊) (Internet café, smoking)</td>
</tr>
</tbody>
</table>

The illustrated record shows that “網咖” and “抽煙” appear in the content and both are 3rd level keywords; since “網咖” belongs to the 2nd level “D1” and “抽煙” belongs to the 2nd level “B2,” this record is classified as sub-category “D1” and “B3.”

After transforming records into keywords, the association analysis and time sequence analysis techniques are applied to discover the co-occurrences of deviant behaviors. Such information could help counselors to ensure that they don’t miss any deviant behaviors and to pre-alert for some deviant behaviors which might happen at almost the same time or in future.

Finally, the last step in Phase II is to apply the discovered knowledge into the existing counseling system. This procedure would function to check memo-type data which is newly entered by counselors so as to immediately classify the data into proper categories. Undoubtedly, it would pre-alert for deviant behavior which a client might be committing in the present or might commit in near future.
The experimental results

In this section, the experimental results of our running example are presented. Since the keyword tree is the basis for our proposed method, it is meticulously verified first. Moreover, the associations and time associations found among deviant behaviors in the counseling records are provided and validated. Finally, the last subsection shows how the proposed framework is incorporated into the existing counseling system and the impacts on users are also measured.

The constructed keyword tree

The keyword tree was constructed by summarizing the keyword analysis results from the Chinese Parser System semi-automatically. Although the list was developed by an experienced officer working over 16 years, we further consulted another domain expert working at the Education Department in Taipei City. The domain expert randomly picked 400 records from the database and classified them manually. Thereafter, the manually classified records were used to compare the correctness of the constructed keyword tree. Finally, adjustments on the keyword tree are made to avoid any biases resulting from keyword selection.

Next, based on the adjusted keyword tree, the domain expert again picked another 200 records to evaluate the quality of the keyword tree by the recall and precision ratio, which have commonly been used in previous studies (Boutell, Luo, Shen & Brown, 2004; Forman, 2003). The definition of recall and precision are defined as follows:

Recall = # of records classified correctly / total # of records that are in the class.
Precision = # of records classified correctly / total # of records that are classified.

By verifying the final version of the semi-constructed keyword tree, the results show that 94.5% of the records were classified correctly. Table 6 demonstrates the detailed comparison results of all recall and precision ratios for each category of the keyword tree.

<table>
<thead>
<tr>
<th>First Level Keywords</th>
<th>No. of records classified correctly by the proposed method</th>
<th>No. of records classified by the domain expert</th>
<th>No. of records classified by the proposed method</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Family Issues)</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td>B (Extroverted Behaviors)</td>
<td>118</td>
<td>120</td>
<td>129</td>
<td>98.3%</td>
<td>91.5%</td>
</tr>
<tr>
<td>C (Deviant Behavior in Learning)</td>
<td>84</td>
<td>89</td>
<td>90</td>
<td>94.4%</td>
<td>93.3%</td>
</tr>
<tr>
<td>D (Deviant Behavior After School)</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>100%</td>
<td>83.3%</td>
</tr>
<tr>
<td>E (Sexual Problems)</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td>F (Introverted Behaviors)</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>97.5%</td>
<td>95.1%</td>
</tr>
<tr>
<td>H (Disciplinary)</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>100%</td>
<td>84.6%</td>
</tr>
<tr>
<td>I (Positive Incentive)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>J (Follow-up Tracking)</td>
<td>31</td>
<td>34</td>
<td>34</td>
<td>91.2%</td>
<td>91.2%</td>
</tr>
<tr>
<td>R (Re-entry)</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Non-deviant Behavior</td>
<td>134</td>
<td>141</td>
<td>136</td>
<td>95%</td>
<td>98.5%</td>
</tr>
</tbody>
</table>

The discovered data associations

General association rules

In Phase II, 32,908 records are analyzed according to the constructed keyword tree. As stated, there are 23,306 records classified as “Others”; in order to examine data records which are not classified properly, we checked whether the associations mined from the “Others” category differ from those mined from the entire collection of data
records. The results show there are no significant differences between the “Others” and overall datasets. It might be concluded that the original taxonomy is not well-suited to classifying deviant behavior and has room for improvement.

The top five association rules mined from all data records are as follows:

Rule 1: Students who “linger about Internet cafés or pool halls” are likely to co-occur with “fighting with people” or “not being focused stably in classes”. (conf. = 96.14%)

Rule 2: Students with “drug abuse” problems are likely to co-occur with “traffic violations”. (conf. = 96.13%)

Rule 3: Students with “emotional barriers” are likely to co-occur with “fighting or hassling with classmates or teachers. (conf. = 20.63%)

Rule 4: Students with “smoking addiction and forgiven by officers” are likely to co-occur with “smoking addiction”. (conf. = 27.68%)

Rule 5: Students with “bad learning attitudes” are likely to co-occur with “cheating in classes”. (conf. = 23.70%)

As readers might notice that in statistics, an association is any relationship between two measured quantities that renders them statistically dependent. The term “association” is often used to emphasize that a relationship being discussed is not necessarily causal. Causality (also referred to as causation) is the relationship between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first event; it is not always true to the relationship of association.

In association rule mining (ARM), our discovered rules are the collections of items that appear to be associated, i.e., they are collections of items that co-occur with unexpected frequencies in the data. Consequently, junior counselors should notice that instead of causality, the discovered deviant behaviors of the rules throughout this paper are to show the possible co-occurrences (associations) under certain deviant behaviors. For example, the interpretation of Rule 1 could be stated as: when a student is reported as “linger about Internet cafés or pool halls,” the information system will pop out “MAYBE CO-OCCUR with “fighting with people” or “not being focused stably in classes” to remind junior counselors to double-check his/her situation.

Association sequence rules

The goal of this step is to mine the sequence associations among deviant behaviors. The discovered rules are mainly used to prevent students from committing other deviant behaviors in near future. Out of our 32,908 counseling records, there are 27,323 individual students with 3,246 of them showing in the database more than once. In total, only 5,585 counseling records can be used for sequence mining. The top two experimental results are provided below.

Rule 1: Students with bad learning attitudes repeatedly show in data records. (conf.: 45.91%)

Rule 2: Students who do not maintain a neat appearance or wear appropriate attire are recorded repeatedly. (conf.: 29.68%)

The above rules show these two kinds of deviant behavior are difficult to remedy. However, comparing to previous subsection, fewer interesting sequences were discovered; one reason might be low frequency counts. Hence, to ensure that the mining of results is more productive, effective approaches for training counselors to track and record sequence data more adequately should be designed. Meanwhile, the sequence associations only from “Other” category are omitted due to no significant differences.

The validation of discovered associations

In order to validate above discovered associations, we invite three experienced domain-experts to verify the agreement of top 9 association and 4 sequence rules. The others are omitted due to low support and confidence values. Table 7 shows the discovered deviant associations reach the average - 4.7 points over five-point Likert scale, which means three domain experts agree with the results in 94% degree. Simultaneously, Table 8 also demonstrates the mined sequence associations reach the average 4.83 points (96.7%). Form above, it can be concluded the discovered associations discovered provide high reliable rules as experienced practitioners have.
Furthermore, the mean absolute error (MAE) of each rule is further examined. Since the confidence of a rule is an estimation of the conditional probability \(P(\text{RHS}|\text{LHS})\), whereas \(\text{RHS}/\text{LHS}\) refers to the right/left hand side of a rule, therefore the MAE is calculated based on the difference between the estimated confidence and the occurred confidence. Thus, we partition whole records into 5-folds along time dimension. Data collected before June 2008 are deemed as training samples to build association rules and the rules’ confidences are calculated as estimated probabilities. Then, data after June 2008 are calculated to testify the differences between the estimated probabilities and occurred probabilities. The MAEs are therefore calculated and as shown high accuracy in Table 7 and 8.

**Table 7. The validation of discovered association rules.**

<table>
<thead>
<tr>
<th>Association Rules</th>
<th>Average agreement</th>
<th>Average MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average agreement = 94% (4.7points) Average MAE = 98.2%</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Students who “linger about Internet cafés or pool halls” are likely to co-occur with “fighting with people” or “not being focused stably in classes”.</td>
<td>93.4%</td>
<td>1.8%</td>
</tr>
<tr>
<td>2. Students with “drug abuse” problems are likely to co-occur with “traffic violations”.</td>
<td>100%</td>
<td>1.4%</td>
</tr>
<tr>
<td>3. Students with “emotional barriers” are likely to co-occur with “fighting or hassling with classmates or teachers”.</td>
<td>100%</td>
<td>1.8%</td>
</tr>
<tr>
<td>4. Students with “smoking addiction and forgiven by officers” are likely to co-occur with “smoking addiction”.</td>
<td>93.4%</td>
<td>2.2%</td>
</tr>
<tr>
<td>5. Students with “bad learning attitudes” are likely to co-occur with “cheating in exams”.</td>
<td>80%</td>
<td>4.2%</td>
</tr>
<tr>
<td>6. Comparing to other students, the ones from “single-parent families” are likely to co-occur with “being less focused on academic learning”.</td>
<td>93.4%</td>
<td>3.3%</td>
</tr>
<tr>
<td>7. Students from “single-parent families” are likely to co-occur with “losing tempers”.</td>
<td>80%</td>
<td>3.8%</td>
</tr>
<tr>
<td>8. Students who “truant frequently” or “with bad learning attitudes” are likely to co-occur with “suspending schooling”.</td>
<td>100%</td>
<td>1.3%</td>
</tr>
<tr>
<td>9. Students with “emotional barriers” are likely to co-occur with “suspending schooling” due to “adaptation problems”.</td>
<td>100%</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

**Table 8. The validation of discovered time-sequence association rules.**

<table>
<thead>
<tr>
<th>Time Sequence Association Rules</th>
<th>Average agreement</th>
<th>Average MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average agreement = 96.7% (4.8points) Average MAE = 98.6%</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Students with “bad learning attitudes” are likely to co-occur with re-offending frequently.</td>
<td>100%</td>
<td>8%</td>
</tr>
<tr>
<td>11. Students without maintaining “neat appearance” or “wearing appropriately” are likely to co-occur with re-offending frequently.</td>
<td>100%</td>
<td>1.1%</td>
</tr>
<tr>
<td>12. Students with smoking additions are likely to co-occur with re-offending frequently.</td>
<td>100%</td>
<td>1.2%</td>
</tr>
<tr>
<td>13. Students who “resume their studies” are likely to co-occur with “re-suspending schooling” due to “bad learning attitudes”.</td>
<td>86.7%</td>
<td>2.6%</td>
</tr>
</tbody>
</table>

With the validations of our findings, junior counselors without many experiences can easily perform their work and surplus efforts and resources can be therefore saved.

**The student counseling system**

**The implementation of new functions**

Finally, Figure 3 shows how the proposed framework could be incorporated into the existing counseling system. Initially, the keyword tree and association rules are mined from existing records and documents. Thereafter, when counselors finish typing counseling details, the important deviant keywords will be identified simultaneously. Of course, the counselor can modify the deviant keywords of the current record by manually adding or deleting improper keywords. Thereafter, these keywords as well as associated rules are retrieved to remind other associated deviant behavior which the client is likely to commit in near future. In this way, the system could advise junior counselors to take adequate steps in time. In following paragraphs, the relevant system interfaces are introduced in four steps.
In the first step, the basic information for the current client is selected according to the keyword tree and also his/her family status is stored (Figure 4). Thereafter, the counseling details are keyed-in and stored as shown in Figure 5.

Next, the counseling system analyzes the current record which the counselor has just entered, and then it outputs the possible categories and deviant keywords as shown in Figure 6. Additionally, the counselor can insert or delete or restore keyword in the current record as needed. Finally, recommendations to pre-alert for possible deviant behavior are provided by the system as shown in Figure 7.

**Figure 3.** A flowchart of the counseling system.

**Figure 4.** The function for storing students’ background information.

**Figure 5.** The function for storing counseling contents.
Figure 6. The function for resolving keywords from the counseling records.

Figure 7. The function for pre-alerting associated deviant behaviors.

The evaluations of the proposed system functions

In this subsection, instead of testing the correctness as previous sections, the impact of usefulness of the proposed method is testified by 30 users from five high schools. First, the adequacy of data records categorized by the keyword tree is evaluated by asking the agreement on users’ perceived usefulness and easiness to use. The average agreement reaches 4.78 points (95.6%) over five-point Likert scale.

Second, the deviant associations supplied by new system functions are also measured by asking users’ perceived usefulness. The average agreement is 4.62 points (92.4%) over 5-point Likert scale. The reason might be the confidence values of some discovered associations are not greater than 50% and such seems to make users’ expectations of the proposed new functions being a little lowered. Hence, additional user trainings could be done before using these new functions.

Conclusion

In this study, we proposed a general framework to mine data associations from memo-type records. A real-life youth counseling case was used to explain all procedures involved in this framework and to verify the adequacy of the proposed approach. Contributions to preventing youth from deviant behaviors are twofold. First, we constructed a keyword tree of deviant behaviors to classify data records. The tree was validated with 94.5% accuracy. Further, from the original database, two thirds of the records were labeled as “Others” and therefore data records were difficult to analyze. Hence, our constructed keyword tree provides a new taxonomy for classifying and analyzing deviant behavior flexibly.

Second, in order to predict the occurrence of deviant behaviors, the associations and time sequences among deviant behaviors were discovered through the proposed framework. With 94% to 96.7% certainty, the discovered associations provide valuable information which has been acknowledged by 3 domain experts; therefore junior counselors could use discovered rules to prevent continued deviant behavior without the help of senior counselors.

Assisting students to desist from committing deviant actions plays an important role in preventing serious crimes.
Counselors need to be very diligent to detect students’ irregular behavior when it appears. Our proposed method incorporates data mining techniques to explore the associations among deviant behaviors. The results are helpful in detecting undesired behavior. In the future, we will continue to incorporate the research results into existing counseling systems so as to improve the effectiveness of school counseling. In addition, we will also test other approaches with regards to establishing the keyword tree.

Acknowledgements
The authors gratefully acknowledge the financial support by National Science Council, Taiwan, R.O.C. under the contract number of NSC 98-2410-H-128-003.

References


Effects of Thinking Style on Design Strategies: Using Bridge Construction Simulation Programs

Chuen-Tsai Sun¹, Dai-Yi Wang²*, and Yu-Yeh Chang³

¹Department of Computer Science, National Chiao Tung University, HsinChu City, Taiwan // ²Department of Science and Communication Engineering, Providence University, Taichung City, Taiwan // ³National Chu-pei Senior High School, Hsinchu County, Taiwan // ctsun@cs.nctu.edu.tw // ophelia.wang@gmail.com // silas@mail.cphs.hcc.edu.tw

*Corresponding author

(Submitted August 19, 2011; Revised December 02, 2011; Accepted December 28, 2011)

ABSTRACT

Computer simulation users can freely control operational factors and simulation results, repeat processes, make changes, and learn from simulation environment feedback. The focus of this paper is on simulation-based design tools and their effects on student learning processes in a group of 101 Taiwanese senior high school students. Participants identified as having executive, legislative, judicial, global or local thinking style tendencies were asked to design bridges using WPBD2007 computer simulation freeware. Design strategies and design tool usage were video recorded and analyzed. Results indicate a positive correlation between judicial thinking style and frequency of substantial change in structure and goal strategy; a negative correlation between local thinking style and frequency of component tool list usage; a negative correlation between use of top-down strategies and frequency of using two types of assistance tools; a positive correlation between the use of bottom-up strategies and frequency of using both tool types; a positive correlation between the use of a substantial change strategy and frequency of using the graphical interface only; and positive correlations between total construction costs and the frequencies of using the substantial change strategy and component list tool.

Keywords
Simulation, Design strategy, Thinking styles, Constructionism

Introduction

Researchers have reported mixed results regarding the effects of instructional technology and daily life technology on learning (see, for example, Greenhow, Robelia, & Hughes, 2009; McLoughlin & Lee, 2007; Spotts, 1999; Ting & Tai, 2010; Zimmerman & Bell, 2008). Various information tools associated with the Web 2.0 model of co-creation and learning are providing teachers with curriculum design and instructional tools such as readily available slides and simulation programs (Ajana & Hartshorne, 2008; Franklin & Van Harmelen, 2007; Greenhow, 2007; McLoughlin & Lee, 2010). These tools can be introduced into learning processes in a natural manner, with or without active promotion by instructors. However, many of them are designed for specific purposes that do not directly match instructional goals or reflect an accurate understanding of learning processes. For instance, search engines may help learners find information, but their use does not necessarily result in learning if students cannot identify information that is accurate or appropriate for their assigned tasks. In other cases, design assistance tools that are not developed for learning purposes require advanced theoretical foundations and/or practical experience. Modifying them to assist learning may produce mixed results—perhaps encouraging experimentation, but possibly increasing reliance on support mechanisms and reducing motivation for creative thinking. Thus, there is a growing body of research on the best uses of multi-faceted information tools in classrooms (Bransford, Vye, Stevens, Kuhl, Schwartz, Bell et al., 2005).

For this project we investigated the effects of simulation-based design assistance tools on learning processes for a group of Taiwanese senior high school freshmen. Taiwanese students must cope with the stress of various entrance examinations, therefore their efforts are primarily focused on attaining high test scores. As a result, quick fixes and useful formulas are deeply rooted in their learning. Emphases on “the only right answer” and “give me the key point, not the reason” have undermined the development of observation, comprehension, and creativity skills in the country’s education system. Simulation tools can be used to address this problem by facilitating exploratory and constructive learning. Instead of forcing students to find “right answers,” simulation programs can give them opportunities to experiment and to develop trial-and-error skills for optimizing the results of their efforts.
Simulation-based design programs can help students perform complex experiments and use acquired knowledge to predict results and modify simulations (Gredler, 2003). Past experience with simulation-based instruction indicates that students with similar knowledge backgrounds have varying preferences for design tools, rules, and strategies, produce a broad range of results, and benefit from a number of concept development processes. We will use Sternberg’s (1997) thinking style theory as part of our effort to explain these differences from a non-ability perspective, based on the assumption that thinking style determines how individuals express their abilities. For our experiment, participants were asked to use a simulation program to design truss bridges. We used the results to determine how thinking style, design behaviors, and design tool preferences affect simulation-based design performance.

**Literature review**

**Simulation and learning**

High-quality simulations give learners more detailed information than printed texts, and their game-like experiences have the secondary advantage of enhancing learning motivation (Cottrell, 2002). Today’s computer simulation products can be used to introduce information on natural phenomena, repetitive processes, and processing procedures—for example, global weather patterns, biology research procedures, chemistry lab experiments, and modeling human behavior (Alessi & Trollip, 2001). In product design, simulations are often applied to test or predict product performance, thus reducing costs. They are also increasingly being used to facilitate trial-and-error experimentation and knowledge construction, with learners adjusting design variables and working with simulation results in real time. Users can fine-tune directions and strategies to close the gap between design results and goal expectations. These processes involve a combination of existing and new knowledge. Alessi’s (2000) depiction of the relationship between simulation fidelity and learning transfer is shown in Figure 1.

![Figure 1. Simulation of fidelity-transfer of learning relationship](image)

In educational contexts, computer simulations can help instructors simplify the operations and output methods of complex systems, and help learners present simulation results that match a range of cognitive abilities. Computer simulations allow learners to perform experiments that cannot be done in the physical world, yet enjoy a sense of achievement similar to that experienced by professional designers (Hwang & Esquembre, 2003). Constructionists support the use of simulations by students to engage in learning through production or design (Papert, 1991), creating sharable products such as computer programs, machines, and games with no cost overruns or risks tied to experimental failure. Such products require students to learn design principles, skills, and strategies.

**Design strategies in simulations**

According to Kafai (1996), most professional design activities concentrate on the final product as the central outcome, whereas in education the central concern is learning process. Thus, students can have effective learning experiences even if they do not create good products. It is important to remember that in simulation design environments that are not intended for learning, students may be misled by a strong product orientation. In some
cases, learners may rely too much on system prompts that are meant to support professional design efficiency, and therefore miss opportunities for independent thinking practice.

The two most common problem-solving strategies associated with classroom design projects are top-down and bottom-up (Kafai, 1996). The first consists of breaking down a problem into meaningful sub-problems, while the second consists of finding solutions from a convergence of ideas during the process of idea implementation. These strategies have been extensively applied to analyses of cognitive control of operational transfer (Schoelles & Gray, 2003), software design procedures (Jeffries, Turner, Polson, & Atwood, 1981), visual search strategies (Wolfe, Butcher, Lee, & Hyle, 2003), and Web searches (Navarro-Prieto, Scaife, & Rogers, 1999). Using computer programming as an example, many researchers believe that a top-down strategy is more effective, since long, complex programs can be divided into a series of smaller, more manageable tasks. However, Kafai (1996) has observed that some students approach problem solving as a “conversation with a situation,” with appropriate designs emerging through a process of implementation—a bottom-up strategy. Kafai also found that the large majority of students in his study used a combination of the two strategies, with various degrees of overlap.

Both strategies have advantages and disadvantages. Simulation environments generally encourage the kinds of bottom-up strategies that are hard to teach and practice in classroom instruction. In contrast, top-down design strategies encourage systematic approaches in which learners plan or choose a structure and then deconstruct it into modules. After establishing details for each module, they must test module integration—an example of a thinking-oriented approach in which all product details must be addressed before completion. Bottom-up design strategies are action-oriented. Simulation software programs consist of numerous components that can be executed individually or in combination, therefore learners can work on understanding the function of each component before searching for feasible mixes. Learners who use bottom-up strategies must be careful to avoid spending so much time on details that they lose sight of the overall goal in the local optimization process. In contrast, learners who use top-down strategies may suffer from limited creativity due to their reliance on conventional thinking models. In this project our concern is whether simulation systems provide learners with excessive guidance for bottom-up design approaches, which are most effective when they balance or supplement top-down approaches.

**Thinking styles**

Sternberg (1994) defines thinking style as an individual’s preferred way of using personal abilities when dealing with life problems. Thinking style represents individual habits rather than intelligence or ability, with ability defined as the extent to which one is capable of accomplishing a certain task, and style defined as the way one prefers getting the task done. In Taiwan, all students must take an entrance examination to attend senior high school; therefore students in any high school class generally share a similar level of academic achievement. In such environments, students’ habits are central to learning performance, and thinking style—as a habit of thought—exerts considerable influence on learning outcomes (Sternberg, 1997). In the context of the present study, it is an important factor influencing the simulation design process.

Sternberg identified 13 styles of mental self-government in the five categories of function, form, level, scope, and leaning. Since our focus is on the effects of thinking style on learning processes and strategies, we will emphasize the function and scope categories. Characteristics of each thinking style are shown in Table 1.

Links between thinking style and design behavior have been examined in studies conducted in countries all over the world, with researchers identifying correlations between such factors as creative ideas and peer feedback (Boden, 2004; Kristensson, Gustafsson, & Archer, 2004; Lin, Liu, & Yuan, 2001; O’Hara & Sternberg, 2001; Wolfradt & Pretz, 2001). Hilton (2002) is among those scholars describing a relationship between thinking style and incentive to implement design products among students in a design department. Our goal is to investigate potential associations between student design processes/learning outcomes and thinking style in a sample of Taiwanese high school freshmen, and the impact of thinking style on differences among students with similar prior knowledge and competencies using simulation software.

According to Sternberg (1994), individuals who use an executive learning style favor fixed and predefined procedures. These learners may rely heavily on instructions provided by simulation software, and approach all tasks in terms of a fixed number of steps. In contrast, learners who follow a legislative thinking style prefer having more
freedom to put together different combinations of components. They are likely to prefer object-oriented design interfaces that support the development of creative ideas. Judicial thinking style is characterized by multiple trials and comparisons, which is a key characteristic of simulation systems. In the scope category, a global thinking style is characterized by greater attention given to the aggregate properties of a design, including asymmetry, aesthetics, and overall costs, while a local thinking style is characterized by greater attention to design weaknesses.

Table 1. Function- and scope-category thinking styles

<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
<th>Behavioral Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>Legislative</td>
<td>Prefer creating their own ways of doing things and deciding for themselves what they will do and how they will do it. Like to create their own rules, and prefer problems that are neither pre-structured nor prefabricated.</td>
</tr>
<tr>
<td></td>
<td>Executive</td>
<td>Prefer pre-structured or prefabricated rules and problems. Like filling in gaps within existing structures rather than creating the structures themselves.</td>
</tr>
<tr>
<td></td>
<td>Judicial</td>
<td>Like to evaluate rules and procedures. Prefer problems involving analyses and evaluations of existing objects and ideas.</td>
</tr>
<tr>
<td>Scope</td>
<td>Global</td>
<td>Prefer dealing with relatively larger and more abstract issues. Tend to ignore or dislike details in favor of “seeing the forest rather than the trees.”</td>
</tr>
<tr>
<td></td>
<td>Local</td>
<td>Like concrete problems rich in detail. More oriented toward pragmatics, and prefer “seeing trees rather than the forest.”</td>
</tr>
</tbody>
</table>

Method

Our motivation was to investigate relationships among thinking styles, design strategies, and design tool preferences for a group of students with similar prior knowledge who were asked to achieve the same goal using a simulation-based bridge design program. We will emphasize the relationship between design tool preference and design strategy, with a secondary emphasis on connections between design outcome and both design strategy and design tool usage. Our research framework is shown in Figure 2.

Data sources

The construction simulation program used in this study was the West Point Bridge Designer 2007 (WPBD2007), developed by Colonel Stephen Ressler of the United States Military Academy at West Point (http://bridgecontest.usma.edu/download2007.htm). WPBD2007 is at the center of an annual international competition for junior high school students, the West Point Bridge Design Contest (http://bridgecontest.usma.edu/). Students from any country can register and send their designs to the Academy, and a small number are invited to
West Point for the final round. This version of the program, which was originally designed to train military engineers, was developed for use by middle school students. The software is regularly updated; the 2007 version was current when this article was being written. WPBD2007 has a simple interface and uses animation, color marks, and graphs to present the results of semi-finished bridge designs. The goal is to design a truss bridge that meets specified loading requirements at the lowest possible cost. As shown in Figure 3, WPBD2007 has an intuitive interface that lets learners draw bridge designs using computer click-and-drag functions. Tool descriptions are given in Table 2.

![WPBD2007 design interface](image)

**Figure 3. WPBD2007 design interface**

<table>
<thead>
<tr>
<th>Tool Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design</td>
<td>Placing bridge components and connection points.</td>
</tr>
<tr>
<td>Component properties</td>
<td>To show the properties of selected components, including loading-length relationships and cost.</td>
</tr>
<tr>
<td>Design test</td>
<td>To test loading capacity.</td>
</tr>
<tr>
<td>Design history</td>
<td>To record all tested designs.</td>
</tr>
<tr>
<td>Status</td>
<td>To show current design cost, simulation test results, and loading test report.</td>
</tr>
<tr>
<td>Component list</td>
<td>Show all component properties and indicate failed components.</td>
</tr>
<tr>
<td>Show tools</td>
<td>Display/hide component list, component number, and design assistance grid.</td>
</tr>
<tr>
<td>File processing</td>
<td>Execute build/open/save/print functions.</td>
</tr>
</tbody>
</table>

WPBD2007 users manipulate design tools to place bridge components and connection points, and then use various tools to check component properties such as load capacity, length, and cost. A design example is shown as Figure 4a. Other tools are available to test preliminary bridge designs. After each loading test, bridge components that are subject to compression are marked in red, and those subject to tension are marked in blue. The difference between an external force and component limit is marked by color intensity—the darker the color, the closer the force to the maximum tolerance limit (Figure 4b). An example of a design that passes the loading test is shown in Figure 5a; one that fails is shown in Figure 5b. The color-marking feature helps users focus on components that need modification. The current bridge design cost is indicated in the status area.
We used a 55-item questionnaire to identify the study participants’ thinking styles—11 items each for executive, legislative, judicial, global and local. Responses were recorded along a 5-point scale (1 = “strongly disagree” to 5 = “strongly agree”). Internal consistency coefficients (α) for the five dimensions were measured as 0.71 for legislative, 0.61 executive, 0.63 judicial, 0.65 global, and 0.45 local. Thinking style tendency refers to the mean score per thinking style obtained by each participant; higher scores indicate stronger tendencies toward a specific style.

Design strategy preference indicates the frequency of using each strategy type over a fixed number of design steps. We considered three design strategies in this project: top-down, bottom-up, and substantial change in design structure and goals. Kafai (1994) discusses the first two strategies in the context of digital games designed for elementary school students. The third strategy, which can be described as “starting from scratch,” cannot be used in conjunction with the other two. Participant design strategy preference scores represent the means of scores given by three raters familiar with engineering design principles. Rater agreement was calculated using Kendall’s coefficient of concordance; coefficients for top-down, bottom-up, and substantial change were .945 (p < .001), .933 (p < .001), and .886 (p < .001) respectively, indicating high levels of agreement.

Design tool preference refers to the frequency of using each design tool over a fixed number of WPBD2007 design steps. The most commonly used design tools are graphical interfaces and component lists. Accordingly, we established three tool usage behavior types: graphical interface only, component list only, and combined use of graphical interface and component list (Figure 6). Design tool preference is directly observable in terms of usage behavior; therefore one member of the research team was responsible for quantifying tool usage behaviors.

Design outcome refers to the cost of a design, as calculated by WPBD2007. To motivate the study participants to achieve this goal, we conducted our experiment as a competition.
Participants

The sample consisted of 101 freshmen students (58 males and 43 females) attending a senior high school in Hsinchu County, Taiwan. Students were identified and recruited by convenience sampling. Taiwanese students do not study mechanics at all in junior high school; they receive some basic instruction in one-dimensional mechanics in their freshman year of senior high school. Overall, the participants had almost no training in bridge design, either direct or indirect; therefore all of their knowledge during this project came from their experience with WPBD2007. There was little likelihood of knowledge from other science classes affecting their performance.

Procedure

The thinking style questionnaire was administered during the first week of the two-week experiment. During the second week, all participants were given 35 minutes of training on the functions of the simulation software and basic knowledge about truss bridges and construction materials. They were given handouts for reference during the 50-minute design activity. Design actions were recorded using screen capture software for later analysis of design strategies and tool preferences.

Results and discussion

Descriptive analysis

Descriptive statistics on thinking styles, design strategies, and tool preferences are presented in Table 3. As shown, the mean scores for participant legislative, executive, judicial, global and local style tendencies were 3.68, 3.34, 3.81, 3.16 and 3.03, respectively. As noted above, all Taiwanese students must take an entrance exam that determines which senior high school they are assigned to. All of the participants had recently taken that entrance examination, which is composed by a large number of multiple-choice questions, and it is possibility that preparation for that test supported a tendency toward a judicial thinking style, which favors analysis and comparison.

As shown in Table 3, the participants used top-down strategies an average of 12.3 times and bottom-up strategies an average of 16.3 times during the first 30 design steps. Both were used much more than the substantial change strategy, with a near-normal distribution of frequency values. In terms of design tool preference, the participants used the graphical interface more frequently, but further analysis of the associated distribution reveals that 22.7% of the participants never used the tool and 35.6% used it exclusively. Usage frequencies were lower for both the component list and component list-plus-graphical interface tools—67.3% and 40.6% of the participants never used this tool or tool combination, respectively.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. value</th>
<th>Std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function of thinking style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legislative</td>
<td>2.00</td>
<td>5.00</td>
<td>3.68</td>
<td>0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>Executive</td>
<td>1.00</td>
<td>4.67</td>
<td>3.34</td>
<td>0.08</td>
<td>0.76</td>
</tr>
<tr>
<td>Judicial</td>
<td>2.60</td>
<td>4.80</td>
<td>3.81</td>
<td>0.05</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Correlation analysis: Thinking style and design strategy/tool preferences

Pearson analyses were performed to identify correlations between thinking styles and design strategies. As shown in Table 4, among the three thinking styles in the functions category, judicial was positively correlated with the frequency of the substantial change of structure and goals strategy. WPBD2007 users can produce several designs quickly, with large amounts of information shown in real time during design tests (e.g., deformations under compression, component properties at various forces, and colors denoting those forces). This platform presents sufficient information and immediate feedback for fine-tuning bridge designs, allowing judicial style learners to react to all possible design ideas, analyze weaknesses and advantages, and revise their designs multiple times. This encourages students to think independently, and complements conventional teaching/learning approaches that are considered more passive.

Design strategy and design tool preference

Correlations between the three design strategies and three design tools were examined using Pearson’s correlation analyses. According to the results shown in Table 5, the usage frequency of top-down strategies was significantly and negatively correlated with the frequencies of using both tools; the usage frequency of bottom-up strategies was significantly and positively correlated with the frequencies of using both tools; and the usage frequency of the substantial change strategy was significantly and negatively correlated with the frequencies of using both tools.
Table 5. Correlations between design strategy and design tool preference (N=101)

<table>
<thead>
<tr>
<th>Design Tool</th>
<th>Graphical Interface</th>
<th>Component List</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design strategy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top-down</td>
<td>0.091</td>
<td>0.100</td>
<td>-0.206*</td>
</tr>
<tr>
<td>Bottom-up</td>
<td>-0.194</td>
<td>-0.048</td>
<td>0.300**</td>
</tr>
<tr>
<td>Substantial change</td>
<td>0.286**</td>
<td>-0.149</td>
<td>-0.260**</td>
</tr>
</tbody>
</table>

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

In WPBD2007, design information is presented in the central graphical interface, and the list of components is shown along the right panel. They present useful information for learners who favor either top-down or bottom-up strategies. According to our video recordings, participants who preferred top-down strategies tended to focus on the graphical interface or configuration of multiple components (i.e., sub-problems) using the components list. When the list of components is launched, the pop-up display blocks part of the design interface (i.e., the component selection or bridge structure change tools). This might not be a problem for professional designers, but could be a barrier for inexperienced learners.

The video recordings also show that participants who preferred bottom-up strategies tended to launch the component list to gather information in the form of statistics and individual component status. In these cases, the component list serves a debugging function and does not interfere with user operations associated with the design interface. Likewise, participants who preferred the substantial change strategy had to use the graphical design interface to revise their designs. Since the pop-up list of components along the right panel might interfere with their design operations, they may have purposefully avoided using it.

Design outcomes and design strategy/tool preferences

Among the study participants, the minimum bridge cost was $144,923.78 and maximum $241,705.72 (mean = $166,144.21; SD = $13,214.65). When graphed, the range of costs produced an approximately normal distribution. Correlations between construction cost and the three design strategies and three tool preferences were examined using Pearson’s analyses. According to the results shown in Table 6, the usage frequency of the substantial change strategy was significantly and positively correlated with total construction cost, and the usage frequency of the component list-only strategy was significantly and negatively correlated with total construction cost.

Table 6. Correlations between bridge construction cost and both design strategy and design tool preference

<table>
<thead>
<tr>
<th>Design Strategy</th>
<th>Design Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-down</td>
<td>Bottom-up</td>
</tr>
<tr>
<td>Substantial change</td>
<td>Graphical interface</td>
</tr>
<tr>
<td>Construction cost</td>
<td>Component list</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construction cost</th>
<th>Top-down</th>
<th>Bottom-up</th>
<th>Substantial change</th>
<th>Graphical interface</th>
<th>Component list</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.124</td>
<td>0.017</td>
<td>0.332**</td>
<td>0.132</td>
<td>-0.215*</td>
<td>0.002</td>
</tr>
</tbody>
</table>

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

Due to the fixed time allotment for the assignment, participants who preferred the substantial change strategy spent most of their time on comparing all possible structures, which reduced the available time for fine-tuning their designs and decreasing construction costs. Our results also suggest that students who made greater use of the component list produced less costly designs. According to our video analysis, participants who made frequent use of the component list were more likely to benefit from information on component specifications and properties, thus creating cheaper and more efficient bridge designs. Additional research is required to determine whether cost information exerts an impact on student learning.

Gender and design strategy, tool preference, and outcomes

As stated earlier, the sample consisted of 58 male and 43 female freshmen students. Results from independent-samples t-tests for thinking style indicate that male students had a significantly stronger tendency than female students toward a judicial thinking style (Mean=3.91 versus Mean=3.67; t(99)=2.501, p<.05), but no other significant differences between genders were noted for the other thinking styles. Furthermore, we failed to find any significant
differences between genders in terms of design strategy, tool preference, or outcome (Table 7). According to these data, gender did not affect simulation learning processes or outcomes in our experiments.

Table 7. Descriptive statistics and t-test results for differences between male and female study participants in terms of thinking style, design strategy, design tool preference, and design outcome.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Male (N=58)</th>
<th>Female (N=43)</th>
<th>t (99)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Thinking Style</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legislative</td>
<td>3.764</td>
<td>.617</td>
<td>3.558</td>
</tr>
<tr>
<td>Executive</td>
<td>3.264</td>
<td>.804</td>
<td>3.442</td>
</tr>
<tr>
<td>Judicial</td>
<td>3.910</td>
<td>.473</td>
<td>3.665</td>
</tr>
<tr>
<td>Global</td>
<td>3.198</td>
<td>.868</td>
<td>3.111</td>
</tr>
<tr>
<td>Local</td>
<td>2.960</td>
<td>.670</td>
<td>3.124</td>
</tr>
<tr>
<td>Design strategy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top-down</td>
<td>11.997</td>
<td>5.961</td>
<td>12.899</td>
</tr>
<tr>
<td>Bottom-up</td>
<td>16.747</td>
<td>6.014</td>
<td>15.891</td>
</tr>
<tr>
<td>Substantial change</td>
<td>1.220</td>
<td>1.675</td>
<td>1.210</td>
</tr>
<tr>
<td>Design tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphical interface</td>
<td>15.740</td>
<td>12.489</td>
<td>19.400</td>
</tr>
<tr>
<td>Component list</td>
<td>5.520</td>
<td>8.103</td>
<td>2.980</td>
</tr>
<tr>
<td>Both</td>
<td>8.740</td>
<td>9.475</td>
<td>7.650</td>
</tr>
<tr>
<td>Construction cost</td>
<td>1.6E5</td>
<td>9.9E3</td>
<td>1.7E5</td>
</tr>
</tbody>
</table>

*p < .05

Conclusion

A well-designed simulation program can help learners predict the results of certain actions, understand processes underly ing observed events, explore the implications of conclusions, evaluate ideas, and develop critical thinking skills (Lunce, 2006). It can also motivate learners to engage in problem solving, hypothesis testing, experiential learning, schema construction, and mental model development (Duffy & Cunningham, 1996). The list of researchers presenting evidence shows that educational simulations with scaffolding, coaching, and feedback features can facilitate the transfer of knowledge from simulation to real-world situations includes Alessi and Trollip (2001), Duffy and Cunningham (1996), and Leemkuil, de Jong, de Hoog and Christoph (2003). In addition to simulation functions, simulation systems also come with numerous assistive tools that utilize the power of modern computer systems for real-time design purposes (e.g., highlighting design weaknesses in various color intensities). These tools are very helpful for experienced designers, but for learners with little or no professional knowledge, they may exert very different effects on learning. For example, local error messages may encourage learners to find quick fixes rather than analyze multiple aspects of their designs or reject original plans.

To identify learner-simulation relationships, we considered learner personalities and thinking style, and observed design strategy and tool preferences. Specifically, our results indicate that students with strong judicial thinking style tendencies were more likely to use system-generated messages to quickly identify design flaws and make adjustments. Note also that the system’s pop-up displays of selected component properties are well-suited to users who tend toward a local thinking style. Sternberg and Spear-Swerling (1996) suggest that teachers should utilize more than one teaching method so as to avoid favoring students with a particular thinking style. This is especially true in learning environments that are heavily reliant on traditional expository teaching methods, in which students are expected to passively take in information. While such a learning model is suitable for executive students, it does not serve the needs of legislative students. The simulation software used in this study represents a new approach, one that allows students to learn according to their preferred thinking styles. The software’s emphasis on more-than-one possible result supports the need of legislative students to establish their own design rules through multiple tests. In addition, since each bridge component has some property limitations (e.g., the relationship between component length and capacity is characterized by fixed tension and stress curves), executive students can refer to specifications when selecting components. During the simulation process, judicial students can use system-generated messages to analyze and evaluate their designs. The needs of global and local thinking style students are met because some of these messages focus on the entire design and others on individual components. In short, the “learning-by-doing”
aspect of the program satisfies the needs of students who have a variety of thinking styles, allowing them to
demonstrate their abilities in their preferred ways.

Participants who used top-down strategies tended to avoid using the two types of available tools at the same time. We
believe the main reason is that they preferred working with a design interface that was neither blocked nor
otherwise disturbed. Participants who preferred bottom-up strategies were much more likely to use both available
tools to facilitate fast access to the results of design adjustments. Using both tools reduced the potential for becoming
mired in local optimization. Participants who favored a substantial change strategy tended to use the graphical design
interface only, a preference that prevented them from identifying core problems. Consequently, they occasionally
had to make major design revisions that added to construction costs.

Simulation software is characterized by a graphical design interface containing a variety of real-time statistics plus a
list of components that support top-down strategies. In addition, parameter adjustment tools and the color marking of
problematic components support bottom-up strategies. According to our findings, design strategy is strongly
associated with design tool preference and design outcome. In addition, simulation software functions considered
appropriate for professionals accustomed to using bottom-up strategies will not cause problems for non-professional
learners, thus making simulation-based learning suitable for classroom instruction.

We used video screen recordings to study participant design actions, including mouse movements, clicks, and drag-
and-drops. Obviously we could not directly observe their responses, mental states, or emotions, thereby limiting our
ability to address factors such as flow, playfulness, professionalism, and enjoyment. We suggest that researchers use
instruments such as eye-trackers to observe other behaviors in order to better capture the benefits of this and other
simulation systems for learning. Furthermore, since our experiment had the form of a competition, there is a need to
determine how learners use simulation platforms and tools in the absence of prize incentives.

Acknowledgements

The authors would like to thank the National Science Council of Republic of China for financially supporting this

References

Gaming, 31*(2), 178-196.


*The Internet and Higher Education, 11*(2), 71-80.


ASEE/IEEE Frontiers in Education Conference, USA, 1*, T3C-1-6. doi: 10.1109/FIE.2002.1158601


Carlsen et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2007* (pp.


A Systematic Understanding of Successful Web Searches in Information-based Tasks

Mingming Zhou

National Institute of Education, Nanyang Technological University, Singapore // mingming.zhou@nie.edu.sg

(Submitted August 23, 2011; Revised September 27, 2011; Accepted January 13, 2012)

ABSTRACT

The purpose of this study is to research how Chinese university students solve information-based problems. With the Search Performance Index as the measure of search success, participants were divided into high, medium and low-performing groups. Based on their web search logs, these three groups were compared along five dimensions of the search process for answers to the assigned tasks: frequency counts of different activities during search process, time allocation on each search activity, search patterns, search query as well as selection of information. The results documented students’ varying abilities to search. The key factor that determines search performance is the effective use of search queries. Successful searchers made use of advanced search options such as extracting relevant and effective key terms from further reading, in contrast with simpler strategies by unsuccessful searchers like adding, removing unimportant words, or using synonyms. High-performing students also showed more monitoring awareness and strategy, such as changing search engines, reformulation of existing search queries as well as backtracking to task questions.

Keywords

Information problem solving, Log analysis, Web search skills

Introduction

In the information era, both social and technological developments have contributed to a situation where information plays a critical role. Computer searching has consequently become the routine behavior for identifying and assessing information to build knowledge. Based on comScore figures, 213 million searches are performed per day within the United States in March 2006. Thus, it is increasingly important that students master the skills to locate, select, evaluate and synthesize information from various sources to meet individual needs especially in their academic studies. Together, these activities constitute a process referred to as information-problem solving (IPS) (Brand-Gruwel, Wopereis, & Vermetten, 2005). This process has been extensively studied, and has resulted in a variety of descriptive and prescriptive models supported by empirical evidence (e.g., Brand-Gruwel, Wopereis, & Walraven, 2009; Ellis, Cox, & Hall, 1993; Kuhlthau, 1993). Across these models, typical elements involved in IPS include identifying learning needs/task interpretation, locating information sources, extracting and organizing relevant information, and synthesizing information from selected sources into cogent, productive uses. In this study, I examined how Chinese university students solve information-based problems and whether the process differs as a function of their performance levels, using computer logs generated while the students performed the task.

Related work

A brief overview of recent literature on IPS clearly reveals that researchers have shifted from investigating search product to search process. Of particular interest to researchers is the comparison between experts and novices in terms of their search strategy use and outcome. Most of these studies have established that experts are more successful in carrying out search tasks than novices. More specifically, experts typically tend to (a) use longer queries (Aula, 2003; Hölscher & Strube, 2000), (b) plan ahead in their searching behavior (Navarro-Prieto, Scaife, & Rogers, 1999), (c) make greater use of Boolean operators and modifiers (Aula & Siirtola, 2005), (d) use more keywords and evaluate the sites based on established criteria (Tabatabai & Luconi, 1998), and (e) pay more attention to the analysis of problems and the evaluation of solutions (Land & Green, 2000). However, in some studies, only small differences were found between the experts’ and novices’ search strategies (Brand-Gruwel et al., 2005) and outcomes (Navarro-Prieto et al., 1999). Another line of research compares search behavior and outcome among domain expert and non-expert. Research consistently shows that domain experts had superior performance over domain novices in terms of efficiency and effectiveness. That is, experts took less time to complete the search tasks and produced a larger number of correct solutions (Lazonder, Biemans, & Worpeis, 2000). They adopted longer and
more complex queries than novices, and used elaborations as a reformulation strategy more often compared with the simple stemming and backtracking modifications used by novices (e.g., Hembrooke, Gay, & Granka, 2005; Zhang, Anghelescu, & Yuan, 2005).

Nonetheless, variation exists in how the level of expertise or web experience is distinguished. Some studies propose that users with a little over one working week (50 hours) of experience are experts (Lazonder et al., 2000); others view final-year doctoral students (Brand-Gruwel et al., 2005), or Internet professionals as experts (Hölscher & Strube, 2000). The label “novice” has been assigned either to individuals with less than five years of computer experience and less than one year of web experience (Jenkins, Corritore, & Wiedenbeck, 2003), psychology freshmen (Brand-Gruwel et al., 2005), or psychology students with one year of web experience (Navarro-Prieto et al., 1999). Thus, the experts in one study could easily be novices in another. The same applies to the definition of domain experts and non-experts. The measurement of subject-domain expertise ranges from having specialized training or having performed research in the field (Jenkins et al., 2003), the achievement score in the subject (Lazonder et al., 2000), answers to fact-based questions (Duggan & Payne, 2008), to self-rated familiarity with the domain (Hembrooke et al., 2005). Aula and Nordhausen (2006) argued that this ambiguity in definition has resulted in difficulties in comparing the results from different studies directly, which does not add much to the understanding of the characteristics of different groups. Also the observed contradictory results from different studies may not actually be contradictory, but simply due to the fact that very different groups of users are being compared.

In addition to the terminology related to search experience or domain expertise, the way the search outcome is measured also varies considerably between studies (Aula & Nordhausen, 2006). Different measures for success are used, such as task completion time, number of tasks solved successfully, number of pages browsed, or accuracy of the answers, which again makes it difficult to compare the results across studies. One shortcoming in using task completion rate alone as a measure of search success is that a searcher can be very slow but eventually find the correct answer, whereas another could find the answer almost immediately yet the quality of the answer is much poorer. As both effectiveness and efficiency are important in determining the success of search, they need to be considered together, not independently as in most of the previous studies.

To address this issue, Aula and Nordhausen (2006) employed the level of experience as one variable predicting search success, which was measured by the task completion time (efficiency) and the number of tasks completed (effectiveness), which were combined together to form a single measure – task completion speed (TCS), to avoid the risk of considering fast but unsuccessful performance as superior to slow but successful. Using this approach, they found that the increase in the years of web use, speed of composing queries, average number of query terms per query and the participants’ evaluation of their search skills were all related to improvement in TCS.

Although Aula and Nordhausen’s (2006) study provided valuable insight into the successful strategies in IPS beyond the previous knowledge of the expert–novice differences, the data was retrieved from observations, think-aloud, and post hoc interview. These methods indeed provide invaluable data concerning the mental activities involved in the process, including the reasons, intentions for the engagement with certain behavior. Yet they are useful in uncovering new or emerging phenomena, rather than testing or confirming what is already known (Young, 2005). Further, given that individuals have a limited short-term memory capacity for talking and attending at the same time, the process of thinking aloud may consume limited attentional resources (Schraw, 2000), and their cognition may differ from what would have been the case had they not verbalized by thinking aloud (Leow, 2002). Thus, task performance of participants who used this method could decline (van den Haak, de Jong, & Schellens, 2003).

**Tracing method in IPS research**

Winne (1982) introduced the term *traces* to describe data that reflect actual learners’ cognitive engagements in learning tasks. Traces are accretion data gathered relatively unobtrusively as learners work on tasks. In contrast to traditional measurement, computer traces can document the dynamic, situated nature of learning, as well as individual event-based differences in activity (MacGregor, 1999). Computer traces are believed to be more accurate than self-reports that require recall of actions and advantageous over concurrent think-aloud protocols, which have been found sometimes to interfere with task performance (van den Haak et al., 2003), and live researcher observations which are constrained by human ability to attend to and record detail (Nesbit & Hadwin, 2006). Empirical data further proved that there is little or no correspondence between the retrospective self-reporting on one
hand and the online log-file measure of students’ cognitive processing on the other (e.g., Sins, van Joolingen, Savelsbergh, & van Hout-Wolters; 2007; Winne & Jamieson-Noel, 2002).

With computer data, information such as search queries, hyperlink URLs, timing of each action and so forth can all be available. The frequency counts, timing and patterns of different study actions can be easily identified and, together with other learner characteristics, contribute to explaining the level of task performance. More importantly, the regulation of behavior during IPS process is made explicitly and available in this approach with high accuracy guaranteed. Computer screen capture tools have been found quite useful which are used to record searchers’ mouse behavior as well as audio input in real time (e.g., Buhi, Daley, Fuhrmann, & Smith, 2009; Tu, Shih, & Tsai, 2008). In this way both students’ search behaviors and the reasons behind their behaviors could be analyzed and understood. Yet the main data analysis method is restricted to labor intensive video data analysis. Another type of computer data collection for online learning is through log files (e.g., Ford, Miller, & Moss, 2005). This format of data allows the use of advanced analysis techniques that could have examined the research questions from a different perspective. While this unobtrusive and accurate record (i.e., computer videos and logs) offers greater details about how students search, it only serves well the analysis of actual search behavior as students’ implicit thoughts and beliefs during the search are not necessarily transferred to explicitly observable online behaviors.

The use of traces in the investigations of students’ online search behavior has gained more and more attention to describe and analyze this dynamic process. According to Jansen and Spink (2006), the studies in this area fall into three categories: (1) those that describe users’ search behavior in an experimental setting, (2) those that use transaction log analysis over a period of time or across a number of users, and (3) those that examine issues related to or affecting web search. This paper focuses on the first line of research. Unfortunately, only a few empirical studies in IPS research up to date fall into this category. Roy and Chi (2003) used computer logs, coupled with field notes, to explore how eighth-grade boys versus girls searched on the web for a task. Each student’s search behavior was diagramed out and a series of six different “search moves” were derived. Results revealed that boys tended to employ a different search pattern from girls and that this variation in search behavior was related to the pattern of performance outcomes. In Leroy, Lally, and Chen’s study (2003), the participants were divided into three groups according to overall actual performance and showed how each group interacted differently with the query optimization feedback. Log analyses showed that this feedback significantly helped low achievers but seemed to hinder high achievers. Again, these studies do not meet the goals of the current research.

In sum, there are studies on skilled versus unskilled searchers’ search strategies, studies adopting more reliable measures of search success, as well as studies using computer logs. However, studies focusing on all these aspects are scarce. The present study aims to meet this gap by using computer logs to investigate how adult learners engage in and regulate their IPS process and whether successful searchers exhibit different search patterns from unsuccessful searchers. The underlying questions of this research are: (1) what types of activities students engage in when solving information problems, (2) how students regulate their IPS process, and (3) whether successful searchers show a different search pattern from unsuccessful searchers.

**Method**

**Participants**

The participants were recruited via advertisements in a Chinese university. Twelve male Chinese university students majoring in computer-related areas (e.g., software engineering, telecommunication, etc.) volunteered to participate in this study (mean age = 20.3). On average, they spent 37.7 hours per week on the Internet. Among different online activities to accomplish academic tasks, searching information was reported to be the most frequent activity (rating of 4 out of 5), followed by downloading relevant material from course website (rating of 3.3 out of 5), and searching materials in the university’s online library (rating of 3 out of 5).

**Apparatus**

During the search tasks, the participants used a PC workstation. All participants used Firefox as the web browser. Google was set up as the default search engine, but the participants were allowed to use any search engine they...
preferred. All searches were performed in Chinese. For the data analysis, a computer tracking tool, SCOOP (Zhou, Xu, Su, & Liu, 2011) was developed which supports web search and collects online behavioral data (traces) without inserting potentially disruptive manipulations into student IPS. The tool records fine-grained traces of user click behaviors (time, duration, and object of each mouse-clicking), students’ navigation paths across the Internet as well as the content students input during the search (e.g., search strings). These actions traced during the student’s IPS session are logged to the second and analyzed to make inferences about student IPS activities.

**Tasks**

Two search tasks were assigned to the participants. The searching topic, how bees make decisions for new home locations, was designed in the way that (1) it was similar to academic tasks students usually encountered in university; (2) participants' prior knowledge would not vary much (the participants generally reported a lack of prior knowledge on the topic); and (3) the questions were sufficiently complicated to avoid successful search by only a few simple attempts (piloted with other undergraduate students with similar backgrounds). The task for the participants was to answer two short answer questions (200-300 words each):

1. How do bees choose where to build their new homes? (close-ended);
2. What do you think are the implications for human life? (open-ended).

Both the task questions and students’ answers were presented in Chinese. The students were allowed to finish the task in their own paces. They were also encouraged to highlight the information they found relevant to the task by using the highlighter in SCOOP.

**Procedure**

The study was conducted in a computer laboratory at the university. After arriving at the laboratory, the researcher informed the participants about the purpose and the procedure of the study. The participants were told that the purpose was “to study their normal information search strategies,” without knowing that their search behavior was tracked at the same time they search. As the aim was to make the search session as typical as possible, it was also emphasized that the students could use any search engines as they wished to complete the tasks. There was no strict time limit for the individual tasks. However, the time limit for the whole search session was limited to one hour. The participants were shown how to log in and make highlights before they start the task. To collect the information they found, the participants were encouraged to highlight the information they found relevant for the task.

**Data analysis**

Four students’ logs were lost or incomplete due to their failure to follow the experiment procedure. Hence, 8 logs constituted the dataset for the analysis below.

**Search performance index (SPI)**

Search Performance Index (SPI) is a measure of the performance in the search tasks, which considers both the efficiency and effectiveness of the search, both the accuracy of the answers (effectiveness) and the total time of task completion (efficiency). The SPI is similar to the task completion speed used by Aula and Nordhausen (2006). However, in their measure, the effectiveness of the search is indicated by the number of tasks completed, whereas the actual score of the answer to each task is employed in this study. Given there are only two tasks in the current study, the use of scores of the answers provides greater variation when interpreting the difference among participants. The SPI measure was calculated with the following formula:

\[
SPI = \frac{Accuracy}{Time} \times 60
\]

In this formula, total time of task completion is entered in minutes. Thus, the quotient (accuracy of answers/task completion time for both tasks) is multiplied by 60 to denote the speed of the task accuracy per hour. The accuracy
of answers was scored between 0 and 5. For the first close-ended question, participants received 5 points if the answer is to the point (i.e., answering what has been asked in the question), elaborative (i.e., including one or more supporting statements or evidences), and logic (i.e., explaining the ways bees select new home locations logically). For the second question, students needed to search web resources, analyze and critically evaluate web materials, and possibly put personal thoughts into answers to complete that task. There were no standardized answers for the second question. The students’ responses of this question were rated on their soundness, richness and organization from 0 to 5 points. In sum, the total score of the two searching tasks were 0–10. Each participant’s answer was reviewed by two faculty members. The final score was the average of the two. Any difference more than three points were resolved through discussions.

The participants were split into three levels based on the distribution of their SPI (M = 10.87, SD = 11.9): high performance (N = 3; M =23.47) medium performance (N = 2; M = 8.29) and low performance (N = 3; M = 0). The three groups were compared along five dimensions: frequency counts of different activities during search process, time allocation on each search activity, search patterns, search query as well as selection of information.

Computer log analysis

Leard and Hadwin (2001) identified four major categories of log file analysis employed in the literature: (a) frequency analysis, (b) patterns of activity, (c) time-based analysis, and (d) content analysis. This guides the analysis of log files in this study.

Frequency analysis, the most prevalent method, involves recording the frequency of specified learner actions, such as accessing a relevant web page. Frequencies are usually treated as continuous variables and analyzed with parametric methods (e.g., analysis of variance or regression analysis). A limitation of frequency data is that it describes actions at one point in time but does not capture relationships across events or the time invested in those event-based actions (Misanchuk & Schwier, 1992). Patterns of activity address that limitation by considering contingent frequencies of specific actions or events. When activities occur in consistent proximity to one another (in a particular sequence), they provide information about learners’ strategic actions. Two common approaches for analyzing patterns of activity are transition matrices and time-based diagrams. The former is most commonly used to track two-event sequences, whereas the latter provides a visual overview of a learner’s entire interaction with the learning environment with each event types represented by symbols graphed on a timeline.

Great potential exists for examining not only the occurrence of events in sequence, but also duration and overlap in those events. The timing of every event available from the logs can be used to infer the duration of time students spend on particular activities, and total or mean time spent studying overall. Inferring duration from log data is important because it cues the researcher to consider what is going on for the learner between logged events and to consider time on task as an important element of engagement (Rouet & Passerault, 1999). To date, techniques for incorporating duration into the graphing of activity patterns have not been adequately explored (Nesbit & Hadwin, 2006). Meanwhile, it becomes important to examine the actual content students develop or work with to evaluate the depth of cognitive processing, to the extent that learning environments afford opportunities to create information objects such as highlighting, notes, etc. In IPS research, this analysis technique is most relevant to the choice of search queries.

Results

Frequency analysis

Frequency counts were calculated for each group on the main searching activities as shown in Figure 1. Consistently, the medium-performing group was more active in most aspects of search process. They visited and revisited more web pages, reviewed the task questions more often and made more highlights than the other two groups. The high-performing group used more search queries, and switched the search engine most frequently. Clearly, attempts with different search queries and search engines were deemed as effective strategies to locate information. The low-performing group consistently scored lowest in most dimensions. They accessed much smaller amount of information, did not review the task questions as much as the other groups and made least highlights during search.
The lack of sufficient engagement with the task could account for their inactiveness to some extent. Interestingly, this group used the same number of search queries on average as the high-performing group and tended to revisit the web pages more often than the high-performing group.

**Figure 1.** Search activity frequencies across groups

**Time-based analysis**

The frequency counts of learner activities provide overall information for a given individual or group by taking the problem-solving process as a whole, but they fail to reveal the total or mean time spent overall or on each component of the search process as a measurement of engagement. This can be solved by time-based analysis. As shown in Figure 2, the medium-performing students spent most time in total, followed by the high-performing and low-performing group. When the total time was broken down, the high-performing group typically allocated more time on searching (including constructing/updating search queries and evaluating web pages from the hit list), whereas the medium-performing group scored highest in reading the content of selected web pages and answering the questions. Surprisingly, they spent least time in searching. The low-performing group consistently spent less time on all stages of search. Within each group, both medium- and low-performing groups spent most time in constructing the answers, followed by reading online information then searching. Yet the high-performing group allocated slightly more time to accessing information than task writing.

**Figure 2.** Time allocation strategy across groups

**Patterns of activity**

Prior research (Branch, 2001; Brand-Gruwel, et al., 2005) repeatedly shows that students lack regulatory skills to successfully solve information-based problems. In this study, there was no instance wherein students located an answer with only one attempt. In other words, students needed to adapt their search strategies when they were unsuccessful during their initial attempts. Based on the log files, five major activities were observed during IPS:
search (submit a search query), read information (open and view a web page), review task instructions (click the web page which contains the task questions), highlight relevant information, and write answers. To better capture the regulatory process, pattern-based analysis was conducted to describe the sequence of activities performed during a specific period of time and presented in a time-based diagram. The analysis only involved the high and low performance group in order to elucidate the difference between successful and unsuccessful searchers to the greatest extent. A time-series sequence chart was created for each participant from these two groups. Levenshtein distance was employed to measure the similarity between pairs of sequences within the same group to search for a most representative pattern in each group. It is defined as the minimum number of edits needed to transform one sequence into the other, with the allowable edit operations being insertion, deletion, or substitution of an event in the sequence (Kruskal, 1983). A sequence with the lowest mean distance with all other sequences compared to within its group will be taken as the “most representative” pattern for that group. This pattern recognition strategy is not ideal in that it does not capture most commonalities across sequences in the same pattern. Yet it serves the purpose well of selecting a sequence that resembles all other sequences in the same group to the largest extent, especially with a rather small group size. In this way, subject 01 (SPI = 24) was identified to represent the high-performing group (Figure 3) and subject 04 (SPI = 0) for the low-performing group (Figure 4).

The two representative patterns differed in several ways. The most striking difference was that the low-performing group started answering the questions much earlier than their counterparts. It could be interpreted in two ways: They found (or they believed they found) the answers in the first few attempts; or they simply wanted to complete the task as soon as possible. Based on their task performance, the first could be due to the incapability of evaluating the relevancy of information correctly, and the second could be the lack of motivation to perform the task as well as they could. Further, different patterns were observed during various stages of IPS. The high-performing subject focused on information-seeking in the initial stage. In order to find the relevant information quickly and accurately, he switched between “search,” “read information” and “access task questions” quite frequently. This can be taken as a monitoring process, as referring back to the task question to fine-tune the search string was an effective way to stay on task. This task analysis action was deemed as a very important facet in self-regulated learning models (e.g.,

![Figure 3](image-url)  
Note. Activity code (y-axis): 1 = search, 2 = read information, 3 = review task questions, 4 = highlight relevant information, 5 = write answers.  
*Figure 3. Representative search pattern in the high-performing group*

![Figure 4](image-url)  
Note. Activity code (y-axis): 1 = search, 2 = read information, 3 = review task questions, 4 = highlight relevant information, 5 = write answers.  
*Figure 4. Representative search pattern in the low-performing group*
Winne & Hadwin, 1998). In this stage, students generate their own perceptions about what the studying task is, and what resources are in place such that they can construct a strategy. The enactment of the strategy occurred in the middle stage in this group, as seen through their construction of the answers by accessing online information. In the final stage, this group chose to highlight relevant information and revisited their answers with a few more new searches. This can be seen as a final check of their answer. This review exercise exhibited their meta-cognitive control. In contrast, the low-performing subject seemed to have experienced a few unsuccessful search attempts at first, as reflected by their frequent revisit of the task questions. Until they found relevant information (they believed), they started working on the answers with no more search in between. The whole process was ended by a few other reading and highlighting, which implied the uncertainty of the subject about the accuracy of the answer.

Content analysis

Another common strategy observed in all participants was to try new search queries. Interestingly, the three groups determined to use a new search string in different ways. The high-performing group on average decided to use an alternative search query on average every 2.04 minutes, the medium-performing group every 1.04 minutes and the low-performing group every 0.80 minutes. It appeared that high-performing searchers would like to access more information from the hit list before rushing to call it a failure and try different search queries. To find out what search strings students used to better understand the reasons for their (in)effectiveness, content analysis was conducted on students’ construction of search queries as well as the information they considered useful. 7 students simply submitted the task questions in their original form to the search engine. The low-performing group generally used this approach in the first attempt, while the high-performing group used it after several unsuccessful attempts. Another common search approach among most participants was to search with multiple keywords extracted from the task question. Students were also asked to highlight information they deemed useful for the task so that researchers can further assess their information synthesis capability. Content analyses of their answers showed that students highly relied on the information they highlighted to construct their answers, yet the high-performing group did bring in their prior knowledge or paraphrase the information they accessed into a logic and meaningful answer. The organization among different bits of extracted information has been observed to be a problem in the low-performing group, such as the lack of transitions, unclear or illogical relationships between sentences, and inappropriate structures. This reflects their lack of capability to synthesize information effectively.

Discussion and conclusion

In this study, Chinese university students’ IPS skills were examined in terms of frequency counts of different activities during search, time allocation on each search activity, search patterns, search query as well as selection of information. The approach of this study agrees with the work of Aula and Nordhausen (2006), which focuses on the level of actual performance in search tasks as an indicator of expertise. SPI can be a feasible and reliable measure for search success, and it is best used when (1) the study consists of one or more tasks; (2) the score of search performance is continuous; and (3) the nature of multiple tasks is similar to be combined into one single measure. In this way, it can be ensured that the observed effective search strategies can be the strategies experts adopt.

In a nutshell, this study documented students’ varying abilities to complete a search task through searching information online. Some successfully accomplished the task, whereas some completely failed. This was largely associated with the different strategies adopted during the process. Consistent with Hölscher and Strube (2000), successful searchers showed more complex behavior upon unsuccessful attempts, such as reformulations of existing queries, changing search engines, as well as backtracking to task questions. The insufficient engagement in the above
activities has led to poor search performance. Interestingly, full engagement with these activities did not result in the best performance either, as evidenced by the medium-performing group. It seemed that reading (and rereading) more information, re-accessing the task questions and making more highlights did help with the search outcome, yet they were not the key. There are three possible reasons. Given that there was far less pronounced difference for the number of search queries used among the three groups, the quality of the queries could affect the search success. Successful searchers made use of advanced search options such as extracting relevant and effective key terms from further reading, in contrast with simpler strategies by unsuccessful searchers like adding, removing unimportant words, or using synonyms. Another reason was the ability of evaluating the effectiveness of a search query. Clearly, high-performing students spent much more time on trying with different search queries. The judgment of the ineffectiveness of a search string stemmed from the glance of the title and synopsis of each hit in the search result page. This capability greatly enhanced the efficiency of searching process and outcome. The third reason was searchers’ monitoring awareness and strategy. The high-performing group tended to formulate a strategy (e.g., find the relevant information by reviewing the task questions), and adapt the strategy whenever necessary (e.g., do a final check by revisiting the answer with a few more new searches), while low-performing students were not skilled at monitoring the process and less active in adapting their search strategies. They tended to be eager to finish the task while being restricted to the limited amount of information.

The results are not generalizable considering the rather small sample size, uniform gender composition as well as the educational background (all from the computer field). Hence, the study is more exploratory, and serves as a preparation of a future larger-scale study. Also, attention needs to be paid to cross-cultural differences in the patterns of online search. For example, European users generated slightly more queries per session and viewed more retrieved pages per query than American users (Spink, Oztutlu, Oztutlu, & Jansen, 2002). Thus, the current findings may not apply to web searchers in other countries. Further, other data-gathering methods can be used, including eye-tracking (Cutrell & Guan, 2007), self-reports (Rieh, 2002), and observation notes (Roy & Chi, 2003). Each technique captures the process from a distinct perspective with benefits and losses. Traditional techniques are still valuable in that they provide capture the mental activities that are not detectable via traces. Log analyses are promising in that they allow researchers to dig into student actual behavior for any ground inferences to be made. Thus, these methods are not orthogonal, but supplementary to one another. And they are expected to be used jointly to paint a fuller picture of student IPS process.

Future work needs to be conducted with larger samples, multiple tasks of different natures or other data analytical techniques. With a different task, strategic students are expected to exhibit different approaches to solve the task. With a large sample, the behavioral traces can be submitted to more advanced statistical analysis, such as regression analysis, structural equation modeling, and other web-mining algorithms for analysis, such as sequential pattern mining techniques to discover information seeking patterns; associate rule mining techniques to find correlations among students’ search terms and the web pages they view; clustering and classification techniques to group students with similar characteristics or into pre-defined categories based on given individual variables or search action sequences. Analysis results will enable us to better assess student IPS skills, as well as identify their difficulties and challenges they experience. The next step is then to explore how educators can use the findings to improve students' acquisition of more efficient IPS strategies. This information will help educators and content developers to develop instructional environments that scaffold web search skills to improve the process and performance.

References


Predictive Effects of Online Peer Feedback Types on Performance Quality

Fu-Yun Yu1* and Chun-Ping Wu2

1Institute of Education, National Cheng Kung University, No. 1, University Road, Tainan City 701, Taiwan // 2Department of Educational Technology, TamKang University, No.151, Yingzhuan Rd., Danshui Dist., New Taipei City 251, Taiwan // fuyun.ncku@gmail.com // cpwu303@gmail.com

*Corresponding author

(Submitted September 01, 2011; Revised November 08, 2011; Accepted December 06, 2011)

ABSTRACT

This study examined the individual and combined predictive effects of two types of feedback (i.e., quantitative ratings and descriptive comments) in online peer-assessment learning systems on the quality of produced work. A total of 233 students participated in the study for six weeks. An online learning system that allows students to contribute to and benefit from the process of question-generation and peer-assessment was adopted. The regression results indicated that quantitative ratings and descriptive comments significantly predicted the quality of produced work (i.e., question-generation performance) both individually and collectively, and descriptive feedback explained more variance in quality of produced work than did quantitative ratings. The empirical significance of this study and suggestions for online learning system development, instructional implementation and future studies are discussed.

Keywords

Online learning system, Peer-assessment, Performance quality, Types of feedback

Introduction

Several theories and empirical studies have suggested that peer-assessment strategies can enhance the quality of designed products (Olson, 1990; Topping, 1998; Yu, 2011; Yu & Wu, 2011). Existing studies on peer-assessment have mostly focused on investigating: (1) learners’ perceptions of gains from feedback given by peers and attitudes towards peer-assessment (e.g., Venables & Summit, 2003; Wen & Tsai, 2006; Wen, Tsai, & Chang, 2006; Yu & Wu, 2011), (2) the validity and reliability of peer-assessment (e.g., Cho & MacArthur, 2010; Falchikov & Goldfinch, 2000; Hughes & Large, 1993; Smith, 1990; van Zundert, Sluijsmans & van Merriënboer, 2010) and (3) effects of different methods and timings of peer-assessment on learning outcomes (e.g., Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Strijbos, Narciss, & Dünnebier, 2010; van Steendam, Rijlaarsdam, Sercu, & van den Bergh, 2010; van den Berg, Admiraal, & Pilot, 2006).

In light of many of the distinct features and capabilities of networked technologies (such as high processing speeds, large amounts of storage space, learner control, multimedia, simultaneity, instantaneity, space-, time- and device-independence, flexibility, interactivity, among others), numerous online peer-assessment systems have been developed and have demonstrated their efficacy over the past decade. While great strides have been made in this area, a few research gaps still exist. Among these, effects of different types of feedback on produced work still require more investigation. Considering that feedback can be expressed in both quantitative and descriptive terms, its individual and combined predective effects are examined in this study. Three research hypotheses are proposed:

1. The averaged quantitative ratings received online from assessors will significantly predict the quality of produced work.
2. The quality of the descriptive comments received online from assessors will significantly predict the quality of produced work.
3. The combined averaged quantitative ratings and the quality of the descriptive comments received online from assessors for the composed questions will significantly predict the quality of produced work.

By assessing the validity of the above questions, we hope to illuminate ways in which peer-assessment activities and environments can be better designed to improve the quality of examined work.
Literature review

Conceptualization of peer-assessment

Peer-assessment can be viewed as a highly effective collaborative learning activity in which learners produce lesson-related artifacts, critically evaluate artifacts produced by their peers and provide them with feedback, and receive feedback from their peers regarding their own produced artifacts (Ammer, 1998; Falchikov & Goldfinch, 2000; Fallows & Chandramohan; 2001; Topping, 1998; van Gennip, Segers, & Tillema, 2010). Through mutual support and assistance, peers with similar learning statuses assist each other to achieve learning goals and to extend both their knowledge bases and skill levels (Topping & Ehly, 2001). Such collaborative relationships not only help to assess the quality of submissions or student performance according to a set of criteria, but also identify areas that may be improved upon and offer suggestions for ways in which the work or performance may be refined (Boud, Cohen, & Sampson, 1999; Topping, 1998; Topping, 2010). Considering that peers tend to be within or near each other’s zone of proximal developments, peers’ comments may be more easily understood by learners than instructors’ (Ammer, 1998; Fallows & Chandramohan, 2001).

Several cognitive processes are mobilized when students are engaged in peer-assessment activities. Assessing the relative quality and merits of the examined work encourages students to engage in critical thinking. In addition, both social and argumentation skills as well as substantial knowledge in the applied area are required to enable comments to be accepted by peers. Also, when observing peers’ work, students are likely to be alerted to problems that may exist in their own work and be prompted to make necessary modifications. On the other hand, when students receive feedback from assessors, the comments provided may cause cognitive conflict and direct students to deal with their existing cognitive defects. Knowledge structuring and re-structuring are cultivated through various cognitive and discursive processes (such as deeper elaboration of materials, self-reflection, comparison, clarification, adjustment, and so on). All of the aforementioned processes, based on cognitive conflict theory, social constructivism, social learning theory and metacognition, should promote the development of critical thinking, knowledge integration and cognitive and metacognitive abilities (Falchikov & Goldfinch, 2000; Topping, 1998; van Gennip et al., 2010; Yu, 2011; Yu & Wu, 2011).

Current research state of peer-assessment

Empirical evidence spanning more than two decades has generally demonstrated peer-assessment’s benefits to better learning and increased cognitive abilities, such as critical thinking, metacognitive strategy, cognitive re-structuring as well as improved comprehension of learned material (Brindley & Scoffield 1998; Falchikov & Goldfinch, 2000; Gatfield, 1999; Hanrahan & Isaacs, 2001; Macpherson, 1999; Purchase, 2000; Topping, 1998; Tsai, Lin & Yuan, 2002; van Gennip et al., 2010). In addition to cognitive gains, peer-assessment creates affective benefits for students. The fact that someone other than teachers will view their work tends to induce students to take charge of and invest more in preparation of their initial work and the subsequent revision of the work. These behaviors are, in essence, a manifestation of increased motivation and sense of responsibility toward learning (Humphrey, Greenan, & McIlveen, 1997).

Recently, to help alleviate the time-consuming, tedious and effort-demanding aspects of using a paper-and-pencil format in a large class setting (such as, collection, preparation and assignment of student work, compilation and calculation of feedback provided by peers and the returning of feedback to individual students, and so on), numerous online peer-assessment systems have been developed to manage peer-assessment (for instance, CAP, NetPeas, Vee heuristic, Web-SPA and SWoRD) (Cho & Schunn, 2007; Sung, Chang, Chiu, & Hou, 2005; Tsai et al., 2001). By using computer and network technologies, such developed systems have demonstrated efficacy for carrying out associated tasks efficiently and effectively (Cho & Schunn, 2007; Davies, 2000; Lu & Bol, 2007; Sung et al., 2005; Trautmann, 2009; Tsai et al., 2001).

While great strides have been made in this area and researchers commonly agree upon the importance of clear and pre-specified criteria for objective peer-assessment (Crane & Winterbottom, 2008; Gielen et al., 2010; Orsmond, Merry, & Reiling, 1996), some research gaps still exist. Recent research has focused on the effects of different types of feedback (e.g., concise feedback versus elaborated feedback, justified feedback versus non-justified feedback).
(Crane & Winterbottom, 2008; Gielen et al., 2010; Nelson & Schunn, 2009; Yang, Badger, & Yu, 2006) and the nature of feedback (Cho & MacArthur, 2010; Gielen et al., 2010; Strijbos et al., 2010; van Gennip et al., 2010).

In light of the fact that feedback can be expressed in both quantitative and descriptive forms, “if and how these different types of feedback provided lead to the improvement of the quality of submitted work” is a pertinent issue that should be pursued. Since some online peer-assessment systems only require assessors to rate the quality of the examined work along a set of rating scales, while others enable assessors to include elaborate feedback for the author’s consideration, their individual and combined predictive effects on the quality of the produced work will help to establish guidelines for better peer-assessment system design and implementation.

Method

Participants and implementation procedures

Two hundred and thirty-three 5th graders from eight classes participated in the study for six consecutive weeks. In view of the beneficial effect of student question-generation on learning (Brown & Walter, 2005; Rosenshine, Meister, & Chapman, 1996; Yu & Liu, 2005; Yu & Liu, 2009; Yu et al., 2005), it was chosen as the learning activity and thus produced questions would be students’ work. Participants were informed that the introduced online question-generation and peer-assessment activity was intended to augment their science learning.

For the duration of the study, once each week, after attending three instructional sessions allocated for science, students participated in a 40-minute learning activity in a computer laboratory. To ensure that all participants possessed the fundamental skills required for the introduced activity, a training session on generating questions and assessing peers’ work (the generated questions in this case) was arranged at the beginning of the study. Considering that true/false and multiple-choice questions are among the most frequently encountered question types in primary schools in Taiwan, these two types of question-generation options were chosen. Each week students were directed to individually generate at least one question each type regarding the instructional content covered that week, as well as to assess at least two questions from a pool of peer-generated questions for each question type.

Online learning system

A learning environment (Question-Authoring and Reasoning Knowledge System, QuARKS) that allows students to contribute to and benefit from the process of constructing questions and receiving feedback from their peers about the composed questions was used (Yu, 2009). Essentially, the question-generation sub-system allows multimedia files to be included as parts of questions, as well as the use of various fonts, text sizes and styles (see Figure 1).

![Figure 1. A screenshot of short-answer question-generation](image)

The peer-assessment sub-system, on the other hand, allows assessors to give evaluative feedback using an online assessment form. Using the form’s built-in criteria (strengths: concise question-stem and options, important
concepts, well-explained notes; weaknesses: unclear question-stems, overly-complicated question-stems, excessively verbose options, multiple correct answers, elusive phrasing and implausible distracters, the left-bottom portion of Figure 2), assessors can assess the overall quality of the generated question on a five-point rating scale (from “far above average” to “far below average”) and rate their recommendation of the question to be included in a drill-and-practice item bank (from “Highly recommend” to “Do not recommend at all”). Also, assessors can add descriptive comments and provide detailed suggestions regarding the question being examined in a designated feedback space (the bottom right portion of Figure 2).

Variables

For the purposes of this study, three variables were included: quantitative ratings of the examined question, the quality of the descriptive comments assessors provided regarding the assessed question item and question-generation performance. Quantitative ratings consisted of two indicators: the overall quality of the produced work (i.e., generated questions) and recommendation for inclusion in the drill-and-practice database. The overall quality and recommendations received from assessors were averaged per question per week throughout the activity.

To ensure objective assessment of students’ work (questions generated and descriptive comments provided), a set of criteria and procedures were developed. The quality of the descriptive comments received from assessors about the composed questions and student question-generation performance was defined against a set of criteria. For peer-assessment, all comments question-authors received with regards to a specific question item were analyzed against a pre-defined scheme and were averaged. The average scores per question per week were then summed up. With reference to Nelson and Schunn’s (2009) study on feedback and Yelon’s “open communication” (1996) instructional principle for providing feedback for students’ performances, the quality of descriptive comments was evaluated on four discrete levels: general comments, specific comments where strengths and weakness are identified, identification of areas for improvement and explicit suggestions for further refinement of questions. Definitions and examples of each of the four assessment levels are listed in Table 1.

Table 1. Four levels of descriptive comments, their definition and examples

<table>
<thead>
<tr>
<th>Levels of descriptive comments</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>General comments</td>
<td>Comments provided only contained built-in comments directly from the system or were general in nature.</td>
<td>✓</td>
</tr>
<tr>
<td>Specific comments where strengths or weakness are identified</td>
<td>The strengths and/or weaknesses of the posed question were clearly identified and explained in the assessor’s own words.</td>
<td>✓</td>
</tr>
<tr>
<td>Identification of areas</td>
<td>Not only were weaknesses</td>
<td>✓</td>
</tr>
</tbody>
</table>
For improvement of the posed question identified, areas for improvement were offered.

Options is good. It would be better if some scenarios could be added. It’s good that the tested concept is important (i.e., noise and its measuring unit). However, it tested trivia facts that no one is expected to memorize (what is the decibel level of noise observed at flight take-off).

Explicit suggestions for further refinement of questions

Explicit suggestions for further refinement of the question were provided. It is nice that you used a character in your question, but your question is not clear enough. Might change it to: Mickey feels the construction site is too noisy. What is most likely the decibel level of noise at the construction site? Also, you may consider adding a picture depicting the surroundings of the construction site.

The scenario and characters in the question stem appeals to the reader. The tested concept (wind instrument) is an important topic to be mastered. However, the options are too simple. People can easily find the answer right. Try to include borderline examples or other less seen instruments (e.g., wooden fish or tambourine) to increase its item difficulty.

In reference to the Torrance creativity index (1974) and King’s question cognitive levels (1992) and questions generated by students, the following criteria were adopted to assess students’ performances in question-generation: fluency, flexibility, elaboration, originality, cognitive level and importance. Each of the indexes was operationally defined to ensure objective assessment. Specifically, each question was graded along the following six dimensions:

- Fluency (0-3) assesses the correctness of wording and punctuation, clarity of meaning and logic, and the relevancy of the constructed question.
- Flexibility (0-2) gauges the interconnectedness between the currently covered topic/unit and prior topics/units and any self-derived examples.
- Elaboration (0-2) assesses the refinement of the questions in terms of creating scenarios for the question, using multi-media files to enhance understanding of the question, the discrimination of alternatives, and so on.
- Originality (0-2) examines the uniqueness of a specific question as compared to those of peers (i.e., innovative methods of formulating questions or content/ideas).
- Cognitive Level (0-3) evaluates the cognitive level demanded of the question author: fact, comprehension, or integration (King, 1992). Comprehension indicates that students used their own words to define or describe learned content, whereas fact indicates that wording in the question is very similar to that used in the instructional content. Integration indicates that a link has been built across topics/units and that explanations have been provided to build connections.
- Importance (0-1) evaluates the importance of the concepts assessed in the constructed question.

To establish inter-rater reliability, ten pieces of student work (questions generated and comments received about the question) created during two out of the six implementation weeks were randomly selected from eight participating classes and evaluated by another independent rater (N = 160). The results of the inter-rater reliability were $r = 0.73, p < 0.01$ and $r = 0.94, p < 0.01$ for question-generation and peer-assessment, respectively. They proved to be satisfactory.

Results

Descriptive statistics and relationships among examined variables

The means and standard deviations of the quality of feedback received on the composed questions (including quantitative peer-ratings and descriptive comments) and students’ question-generation performance are listed in Table 2.
### Table 2. Descriptive statistics and correlations between variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Quantitative ratings</th>
<th>Descriptive comments</th>
<th>Produced Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative ratings</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive comments</td>
<td>0.1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Produced work</td>
<td>0.28**</td>
<td>0.37**</td>
<td>1</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.45 (0.68)</td>
<td>6.70 (3.60)</td>
<td>32.11 (13.56)</td>
</tr>
</tbody>
</table>

Note. N = 233, * p < 0.05, ** p < 0.01

### The predictive effect of quantitative ratings on the quality of produced work

The peer-assessment activity following the question-generation activity allowed question-assessors to rate its quality and to make a recommendation on the assessed question. A preliminary analysis using Pearson correlations was conducted to determine the relatedness of the quantitative ratings and question-generation performance. The quantitative ratings were significantly correlated with question-generation performance ($r = 0.28, p < 0.01$); therefore, this variable was included in the regression analysis. The regression results presented in Table 3 indicates that the quantitative ratings significantly predict question-generation performance ($\beta = 0.28, p < 0.01$).

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>SEB</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>12.72</td>
<td>4.46</td>
<td></td>
</tr>
<tr>
<td>Quantitative ratings</td>
<td>5.60</td>
<td>1.27</td>
<td>0.28**</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>19.59**</td>
<td></td>
</tr>
</tbody>
</table>

Note. a. Predictor: (Constant), Quantitative ratings  
   b. Dependent variable: Question-generation performance  
   c. * p < 0.05, ** p < 0.01

### The predictive effect of the quality of descriptive comments on the quality of produced work

The same preliminary analysis using Pearson correlations was conducted to determine the relatedness of the quality of descriptive comments and question-generation performance. The quality of descriptive comments was found to be significantly correlated with question-generation performance ($r = 0.37, p < 0.01$); therefore, this variable was included in the regression analysis. The regression results presented in Table 4 indicate that the quality of descriptive comments significantly predicts question-generation performance ($\beta = 0.37, p < 0.01$).

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>SEB</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>22.80</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td>Quality of descriptive comments</td>
<td>1.39</td>
<td>0.23</td>
<td>0.37**</td>
</tr>
<tr>
<td>R square</td>
<td></td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>36.48**</td>
<td></td>
</tr>
</tbody>
</table>

Note. a. Predictor: (Constant), Quality of descriptive comments  
   b. Dependent variable: Question-generation performance  
   c. * p < 0.05, ** p < 0.01

### The combined predictive effect of quantitative ratings and quality of descriptive comments received on the quality of produced work

To avoid multicollinearity, a preliminary analysis using Pearson correlations was conducted to determine the relatedness of the quantitative ratings and the quality of descriptive comments. The quantitative ratings were not
correlated with the quality of descriptive comments \((r = 0.1, p = 0.13)\); therefore, these two variables could both be included in the multiple regression analysis. As shown in Table 5, the quality of descriptive comments significantly predicted a significant proportion of variance on students’ question-generation performance \((R^2 = 0.14, F = 36.48, p < 0.01)\). Adding the variable of quantitative ratings significantly enhanced the R-square \((R^2\text{ change} = 0.06, F = 16.98, p < 0.01)\); therefore, it was determined that together the qualities of descriptive comments and quantitative ratings can significantly predict question-generation performance \(\beta_{\text{qual}} = 0.35, p < 0.01; \beta_{\text{quan}} = 0.25, p < 0.01\), respectively.

### Table 5. Multiple regression analyses of the quality of feedback’s ability to predict the quality of produced work

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 B</th>
<th>Model 2 B</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>22.80</td>
<td>6.44</td>
<td>6.44</td>
</tr>
<tr>
<td>Quality of descriptive comments</td>
<td>1.39</td>
<td>1.30</td>
<td>0.37**</td>
</tr>
<tr>
<td>Quantitative ratings</td>
<td>4.91</td>
<td>4.91</td>
<td>0.25**</td>
</tr>
</tbody>
</table>

| R-square                       | 0.14        | 0.20        |
| F for change in R-square       | 36.48**     | 16.98**     |

Note. a. Dependent variable: Question-generation performance  
\(b. * p < 0.05, **p < 0.01\)

### Discussion and conclusions

Numerous online learning systems have been developed for learners to interact with their peers online to improve the produced work through peer-assessment. This study explored whether different types of feedback received from peers contributed to the quality of produced work (in this specific case, question-generation performance). Moving beyond the frequently adopted research paradigm that examines assesses’ subjective attitudes toward peer-assessment (Topping, 2010), objective measures were used to examine whether any predictive relationships existed between examined variables.

This study found that quantitative ratings and quality of descriptive comments that question-authors received from peers contributed to question-generation performance both individually and together. Specifically, question-authors who received higher quantitative ratings tended to demonstrate better performance composing questions. Also, better quality descriptive feedback received for composed questions lead to better performance on question-generation tasks. Furthermore, question-authors who received higher quality descriptive feedback together with higher ratings on their questions tended to have better question-generation performances.

The obtained findings have important empirical significance and implications for online system development and instructional implementation. Quantitative ratings and descriptive comments are both forms of feedback that can be easily offered for the author’s consideration during online peer-assessment; however, their effects on the product quality have not yet been substantiated. The present study, to the best of the researchers’ knowledge, evidenced the individual and combined effects of both types of feedback for the first time.

Based on the findings of this study, several suggestions are provided. First, in view of the significant combined predictive effects of quantitative ratings and descriptive feedback, developers of online peer-assessment systems are advised to include both designs in the system to achieve optimum performance. This advice is particularly crucial for online systems that currently incorporate only quantitative ratings for feedback to authors.

Furthermore, this study found that the descriptive feedback variable explained more variance in question-generation performance than quantitative rating, so the importance of providing question-authors with descriptive feedback should not be ignored, especially with regard to those who have received relatively lower quantitative ratings from their peers. As indicated, question-authors who received higher quantitative ratings were found in this study to demonstrate better performance composing questions. Nonetheless, looking at the flip side of the result leads to the conclusion that question-authors who received lower quantitative ratings tended to demonstrate lower question composing performance. As the degree of explicitness and specifics are different between quantitative ratings and descriptive comments, it will understandably be hard for authors who receive lower ratings to improve their performances without elaborate explanations or detailed recommendations from assessors. Providing descriptive
comments to authors should be emphasized to assessors when a lower than average quality of produced work is observed.

Moreover, as has been found, providing a summary of the performance with the locations of specific problems and solutions or explanations is associated with increased understanding (Nelson & Schunn, 2009). Training sessions by instructors about providing quality descriptive comments are essential to maximize the effect of peer-assessment.

Suggested topics for future studies are as follows. As mentioned, the effects of different types of feedback have been the focus of recent research on peer-assessment. While studies along these lines help to provide empirical evidence and build guidelines for constructive peer-assessment learning environments, some areas require further investigation. First, while this study supported the individual and collective predictive effects of quantitative ratings and descriptive comments on the quality of produced work, since that the two different types of feedback demand different amounts of effort from participants (quantitative ratings presumably require less mental effort and time than descriptive comments) and differ greatly in terms of the extent of explicitness of provided information (with descriptive comments providing more specific suggestions and thus presumably offering more support than quantitative ratings), an investigation into the effects of different types of feedback on participation/engagement and interaction cycles, students preference and overall satisfaction, and so on, using an experimental research design method, would have pedagogical and research significance.

Finally, previous research has indicated that the impacts of different types of descriptive feedback (e.g., elaborate and specific versus concise and general) on performance and produced work varied with the competency levels of students involved in the peer-assessment activities. For instance, Strijbos and his colleagues’ study (2010) indicated that concise general feedback from the low-competence assessors resulted in positive learning gains for feedback receivers. On the other hand, Gielen et al (2010) found that high-competency assesses seem to rely less on elaborative specific feedback. With these research findings in mind, further investigation into how the competency level of assessors and assessees interact with types of feedback and influence performance would be practically and empirically relevant.

Acknowledgement

This paper was funded by research grants from the National Science Council, Taiwan, ROC (NSC 96-2520-S-006-002-MY3; NSC 99-2511-S-006-015-MY3).

References


Incorporating Usability Criteria into the Development of Animated Hierarchical Maps

Yu-Cheng Shih¹, Pei-Ren Huang¹ and Sherry Y. Chen¹,²*
¹Graduate Institute of Network Learning Technology, National Central University, Taiwan // ²Department of Information Systems and Computing, Brunel University, United Kingdom // tearstains@cl.ncu.edu.tw // davidfirsttw@cl.ncu.edu.tw // sherry@cl.ncu.edu.tw

* Corresponding author

(Submitted September 01, 2011; Revised January 12, 2012; Accepted March 14, 2012)

ABSTRACT

Nowadays, Web-based learning systems have become popular because they can provide multiple tools, among which hierarchical maps are widely used to support teaching and learning. However, traditional hierarchical maps may let learners easily get lost within large information space. This study proposes an animated hierarchical map to address this problem. To consider the effects of different instructional approaches, we also conducted an empirical study to examine how learners in collaborative learning and those in individual learning reacted differently to the animated hierarchical map and traditional hierarchical map. A questionnaire, which was designed based on Nielsen’s heuristics, was applied to examine learners’ reactions to the proposed animated hierarchical map. The results show that the proposed animated hierarchical map is generally better than the traditional hierarchical map but the effects on individual learning is limited.

Keywords
Hierarchical map, Nielsen’s heuristics, usability, web-based learning

Introduction

Web-Based learning (WBL), which is showing explosive growth (Song, Wang, & Liu, 2011), is widely employed by worldwide corporations and universities (ChanLin, 2010). This is probably because WBL provides high flexibility (Chen & Macredie, 2010). For instance, the WBL allows learners to use multiple navigation tools to develop their own learning strategies. Among various navigation tools, hierarchical maps (HMs) are popular for learners because a graphic representation is used to give an overall picture (Nilsson & Mayer, 2002).

Several studies indicated that such an overall picture is beneficial for learners to build the global understanding of the subject content (Nielsen, 2000). It is probably due to the fact that such an overall picture can help learners construct logical relationships between each topic. In brief, learners can greatly benefit from the HM (Amadieu et al., 2010). However, such benefits only appear when the HM is well designed. Conversely, a massive HM may hinder learners to reach their learning objectives. For instance, learners may find it difficult to discover relevant issues if information is not systematically displayed in the HM. To this end, the design of the HMs should carefully consider usability, which can facilitate users’ interactions with systems and effectively reach the purposes of the systems (Graniča & Ćukušić, 2011).

In addition to taking into account usability, there is also a need to consider the requirements of different instructional approaches. Recently, WBL is not only applied to support individual learning, but also is employed to promote collaborative learning. Regarding individual learning, a learner develops learning strategies on his/her own (Johnson et al., 2010). Regarding collaborative learning, learners work together to develop their learning strategies so they can exchange ideas, perspectives, and arguments with each other (Pargman and Wærn, 2003). The differences between these two approaches lie within the fact that individual learning is the one in which learners work individually to reach their own objectives while collaborative learning is an instruction method in which more than an individual work together to reach their common objectives (Leidner & Fuller, 1997). In other words, individual learning and collaborative learning are different approaches so they have different requirements. Thus, there is also a need to consider these two approaches when we propose a new HM for WBL.

To this end, the study presented in this paper not only incorporates usability into the development of an animated HM, but also examines whether the animated HM can accommodate the requirements of collaborative learning and individual learning. It is our ultimate goal that such an animated HM can work as a benchmark for the improvement
of the design of WBL systems and other Web-based applications. By doing so, the performance and satisfaction of users who uses these applications can be enhanced.

Theoretical background

WBL systems offer multiple tools to help learners locate information effectively, including main menu, keyword search, hierarchical map, and alphabetical index. These navigation tools serve different purposes. A main menu is used to display a list of relate topics in a logical way. A keyword search supports learners to locate relevant information for a particular concept. A hierarchical map uses a graph to illustrate relationships among various topics. An alphabetic index is used to list all of topics in an alphabetic order (Khalifa & Kwok, 1999; Nilsson & Mayer, 2002; Chen & Macredie, 2004). Among these tools, the HM is popular in the WBL systems. This is due to the fact that the HM visualizes logical relationships among various topics with a graph (Majid et al., 2006). As indicated by Levin, Anglin and Carney (1987), combining texts with graphics is a useful instructional means, which can make verbal description more comprehensible and facilitate learners to understand a complex subject matter. In addition, the HM can present whole information within a single graph. As showed in the study by Danielson (2002), the HM can help learners get a global view so that learners can easily build logical relationships between each concept. Furthermore, previous studies found that the HM could reduce learners’ disorientation problems. For instance, Amadieu et al. (2009) found that HMs can help learners avoid getting lost because learners can easily identify their current status and reach their targets effectively.

Due to these benefits, several studies attempt to improve the development of HMs with advanced learning technologies, among which computer visualization and animation have been extensively used (Null & Rao, 2005; Holliday, 2003). Regarding computer visualization, an early study by Johnson and Shneiderman (1991) used an interactive visualization method to show large hierarchies within a limited space. Regarding animation, Dicheva, Dichev and Wang (2005) designed an interactive graphical user interface that combines a hierarchical layout with an animated view, where zoom-able views and click-able topics were provided. Furthermore, a recent study by Nesbit and Adesope (2011) produced an animated concept map in the form of a network diagram in which nodes and links are sequentially added or modified. These technologies bring a lot of improvements to HMs so we also attempted to use these technologies in our proposed animated HM.

However, the aforementioned studies ignore the importance of usability in the development of HMs. Usability is essential for the design of software development (Juristo et al., 2007) and that there is a close relationship between Web-based service quality and usability of Web-based information systems (Oztekina, Nikovb and Zaimc, 2009). As indicated by Van den Haak et al. (2004), usability is useful to improve information systems. They used usability to enhance online library catalogues. Moreover, Karahoca et al. (2010) used usability to improve the emergency department (ED) software prototypes developed for Tablet personal computers to keep electronic health records of patients errorless and accessible through mobile technologies. Recently, Harvey (2011) used usability to improve toolkit for In-Vehicle Information Systems (IVIIs). Additionally, a framework is presented to guide designers through defining usability criteria for an evaluation, selecting appropriate evaluation methods and applying those methods.

The aforesaid studies demonstrate the importance of usability. Due to such importance, several studies also proposed criteria for the assessment of usability. Among various criteria, five criteria proposed by Nielsen (1993), i.e., “Efficient to use,” “Easy to learn,” “Few errors,” “Easy to remember” and “Pleasant to use,” are widely employed to assess usability. For instance, Yeung and Law (2004) applied these five criteria to improve the usability of hotel websites in Hong Kong. On the other hand, these five usability criteria were also included in the criteria of other studies. For example, Campbell (2003) indicated that usability criteria should follow by easy to learn, easy to use, easy to remember, error tolerant, subjectively pleasing. Another study by Brinck (2002) claimed that usability criteria should consider correct functionality, efficient to use, easy to learn, easy to remember, error tolerant, subjectively pleasing. An early study by Wixon (1997) claimed that usability criteria should include learnability, efficiency, memorability, satisfaction, flexibility, first impressions, advanced feature usage, resolvability. Because of the significance of these five criteria, we not only incorporate usability into the development of our proposed animated HM, but also use these five criteria to assess the usability of this HM. The novelty of our study is due to this approach, which is different from the aforementioned studies.

343
In addition to this novelty, this study also considers the requirements of different instructional approaches because many new instructional approaches have appeared in current educational settings in the past ten years. In particular, collaborative learning is widely used to support teaching and learning. For example, Tolmie et al. (2010) used collaborative learning to supported learners in science training, in which learners are grouped to collaboratively work in the context of two science topics, i.e., forces and evaporation. Another study by Prichard et al. (2006) used collaborative learning to supported learners in team-skills training, in which training objectives include problem solving, planning, decision making, setting objectives, time management, agreeing roles, creating a group environment and cooperation. Moreover, the study by Mukama (2010) used collaborative learning to support university students to develop hands-on computer skills and knowledge in small task-based groups.

Conversely, individual learning is different from collaborative learning. Gokhale (1995) clarified the differences between these two approaches. He thought that collaborative learning is an instruction method in which students work in groups toward a common academic goal while individual learning is the one in which students work individually at their own level and move toward an academic goal. These two approaches also have different advantages. Collaborative learning can lead to deep information processing (Kirschner et al., 2009a) and stimulate and enable learners to engage in activities that are valuable for learning (Kirschner et al., 2009b). In contrast, individual learning makes learners autonomously cover all aspects of a learning task and exercise argumentative learning activities at their own pace (Weinberger et al, 2010). Additionally, learners have to mentally interrelate multiple external representations to understand the learning material and the underlying concepts by themselves (Bodemer, 2011). Because of such differences between collaborative learning and individual learning, we also conduct an empirical study to investigate whether the proposed animated HM can accommodate the needs of students who use collaborative learning and those who use individual learning.

Methodology design

The aims of this study are two-fold: (a) to propose an animated HM for large information space and (b) to assess the proposed animated HM by conducting empirical research. As described in the previous section, there is a need to consider different instructional approaches. Thus, the empirical work not only compares the effects of the traditional HM and the proposed animated HM, but also investigates how learners with collaborative learning and those with individual learning react differently to the traditional HM and the proposed animated HM. The details are described in subsections below.

Design of an animated hierarchical map

To achieve the first aim, we initially assess the usability of existing HMs based on five main criteria proposed by Nielsen, i.e., “Efficient to use,” “Easy to learn,” “Few errors,” “Easy to remember” and “Pleasant to use.” The results of the assessment show that the existing HMs have several shortcomings (Table 1). To address these shortcomings, we explore some potential solutions, which are described in Table 1.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Objectives</th>
<th>Existing Shortcomings</th>
<th>Proposed Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy to remember</td>
<td>Learners can still recall its general structure</td>
<td>Massive information is presented so it is difficult to remember where to locate information.</td>
<td>The content is divided into several main sections, each of which includes a number of sub-sections. The subsections can be hidden.</td>
</tr>
<tr>
<td></td>
<td>though they do not use a system for a while.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy to learn</td>
<td>Learners can quickly understand how to explore</td>
<td>Learners need to keep using a scrolling bar to seek additional information.</td>
<td>Information is presented within one page so there is no need to use a scrolling bar.</td>
</tr>
<tr>
<td></td>
<td>the subject content.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficient to use</td>
<td>Learners can quickly orient themselves and</td>
<td>There is a lack of clear visual cues so learners feel difficult to know</td>
<td>Additional visual cues are used to help learners identify their current location.</td>
</tr>
<tr>
<td></td>
<td>identify a right direction to achieve learning.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Regarding the shortcomings of existing HMs, massive information is presented (Figure 1) so learners may easily get lost. To address this problem, we use Adobe Flash to develop an animated HM (Figure 2 and Figure 3) based on the solutions described in Table 1. The animated HM show eight main sections and then learners can use left and right arrow keys to choose a particular main section, of which subsections are displayed above the main section chosen. In
other words, other sub-sections are hidden. By doing so, the animated HM presents a clear structure, which can be easy to remember. Due to the fact that the screen only shows eight main sections and the subsections of the main section selected, all of the subject content can be displayed within one page. Thus, there is no need to use scrolling bar so that students may feel it is easy to learn how to use the animated HM.

Figure 3. The design solutions of the proposed animated hierarchical map

Furthermore to the abovementioned left and right arrow keys, another arrow key placed in the middle side is applied to point out the main section chosen so that students can clearly identify their current location, which, in turn, they feel efficient to use the animated HM. In addition to the chosen main section, the selected subsections are also highlighted with yellow lines so that students can correctly select relevant subsections and make few errors. In summary, arrow keys are employed to switch the HM with animations, which students may feel pleasant to use.

Empirical work

To achieve the second aim, an empirical work was conducted and the detailed design of the empirical work is described below.

Participants

30 undergraduate and graduate students voluntarily took part in our empirical work. To recruit these participants, a request was issued to students in lectures, and further by email, which indicated that all participants had to have the basic computer and Internet skills to access a WBL system but they should not have taken any courses related to the subject content of the WBL system before they participated in our study.
Web-based learning system

In this empirical work, the WBL system presents the principles of “Interaction Design” and includes eight sections. Either the animated HM or the traditional HM is used as an entry point to go into the main content of each section. The screen design consists of (a) a title bar used to present the subject of the WBL system, (b) a MAP button placed in the left hand side and corresponded the animated HM or the traditional HM to help learners locate information from the WBL system, and (c) the main content of the WBL system (Figure 4).

![Interaction Design](image)

*Figure 4. The Web-based learning system*

Design of a questionnaire

A questionnaire was used to identify whether learners were satisfied with the proposed animated HM because it is useful to investigate learners’ perception (Kinshuk, 1996). The questionnaire was designed with open questions, which can allow participants to express their opinions with their own words. The design rationale of the questionnaire was based on the aforementioned criteria listed in Table 1. To this end, seven open-questions are included in the questionnaire. The examples of the questions are presented in Table 2.

<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How do you think about the efficiency of these two types of hierarchical map. More specifically, which map can let you quickly locate relevant information? Please also give the reasons why you think that this map is more efficient.</td>
</tr>
<tr>
<td>2</td>
<td>Please describe your experience in using these two types of hierarchical map. In particular, please identify which map you feel easy to use and give the reasons for your choice.</td>
</tr>
<tr>
<td>3</td>
<td>Please describe your perception to the design layout of these two types of hierarchical map. In particular, please identify which map you feel pleasant to use and give the reasons for your choice.</td>
</tr>
</tbody>
</table>

Post-test

As described in the Methodology Section, all of the participants did not have any prior knowledge of the subject content so only a post-test was applied to assess how much they have learnt from the WBL systems, instead of using both of the pre-test and the post-test. The post-test was presented in a computer-based format and included 20 multiple-choice questions. Each question included three different answers and an “I don’t know” option but there was only one right answer. The questions covered all eight sections of the WBL system from basic concepts to advanced topics and the answers could be obtained by using both of the traditional HM and the animated HM. Students were allotted 20 minutes to take the post-test and were not allowed to examine the content presented in the system when they took the post-test.

Experimental procedures

The participants were divided into two types of learning settings, i.e., collaborative learning and individual learning, in a random way. For each setting, all of the participants need to use both the traditional HM and the animated HM.
By doing so, we could examine the effects of these two types of maps on individual learning and collaborative learning.

Regardless of the traditional HM and the animated HM, the participants were also required to complete some practical tasks when they interacted with the WBL system. The tasks included 20 factual questions, which focused on a single concept and there was only one standard answer for the question. The participants needed to find answers for these questions from the WBL system. More specifically, the purpose of performing these tasks was to offer the opportunities of experiencing interface features provided by the traditional HM and the proposed animated HM. Subsequently, they had to take the post-test, which was used to assess how much they had learnt from the WBL system. In other words, the post-test scores were used to identify their learning performance. Finally, the participants needed to fill out the questionnaire to express their perception to the design of the proposed animated HM and the traditional HM.

Data analysis

Two independent variables are included this study. One is related to the design of HMs (i.e., the animated HM and the traditional HM), and the other one is concerned with learning settings (i.e., collaborative learning and individual learning). There are also two dependent variables. One is learners’ performance, which is assessed by their post-test scores. The other one is learners’ perception, which is identified by their responses to the questionnaire.

Results and discussions

Learning performance

Figure 5 presents the learning performance of students using the traditional HM (mean = 5; standard deviation = 1.49) and the proposed HM (mean = 5.17; standard deviation = 1.40) in individual learning. Figure 6 presents the learning performance of students using the traditional HM (mean = 5.17; standard deviation = 0.75) and the proposed HM (mean = 7.17; standard deviation = 1.83) in collaborative learning. The result from the Independent t-test shows that there is no significant difference between the students using the traditional HM and those using the proposed HM in individual learning (p > .05). It implies that the two maps have similar effects on individual learning. On the other hand, the result from the Independent t-test shows that there is significant difference between the students using the traditional HM and those using the proposed HM in collaborative learning (p < .05). More specifically, students who used the proposed animated HM could get better learning performance than those who used the traditional HM in collaborative learning. This finding suggests that the proposed animated HM is useful to enhance students’ learning performance in collaborative learning.
In summary, the proposed animated HM is not very helpful for individual learning but it is beneficial to enhance students’ learning performance in collaborative learning. The difference between individual learning and collaborative learning lies within the fact that each individual builds his/her learning strategies by himself/herself in the former while individuals work together to build their learning strategies in the latter. Thus, social exchange among various individuals may be a key factor for collaborative learning (Schnotz, Bockheler & Grzondziel, 1999). The above mentioned results suggest that the proposed animated HM may be a useful social exchange tool for students.

Learning perception

Table 3 presents the strengths and weaknesses of the traditional HM and the proposed animated HM for individual learning and collaborative learning based on learners’ responses to the questionnaire.

<table>
<thead>
<tr>
<th>Table 3. The strengths and weaknesses of the traditional hierarchical map and the proposed animated hierarchical map</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traditional hierarchical Map</strong></td>
</tr>
<tr>
<td><strong>Strengths</strong></td>
</tr>
<tr>
<td><strong>Weaknesses</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Proposed animated hierarchical Map</td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Individual learning**

Most of learners in individual learning think that the traditional HM can present the complete information of the subject content at a time so learners can easily get an overall picture at an early stage. It may be because the traditional HM utilizes a tree-based structure, which allows students to easily grasp an overall picture. Conversely, the traditional HM provides massive information. Such massive information may be irrelevant to learners’ tasks so they may experience cognitive overload, which may hinder them from reaching their learning goals (Chen et al., 2009). Additionally, learners are provided with too many choices so learners may feel difficult to select right information for their tasks.

Regarding the proposed animated HM, most of learners in individual learning think that the proposed animated HM is useful to examine a single topic deeply. It may be because the subject content is well classified and presented in a systematic way. This is also the reason why the learners value the proposed animated HM. Moreover, the proposed animated HM can not only present sufficient information within one page but also the size is not as big as the traditional HM. Consequently, they think that the proposed animated HM offers a friendly interface and feel easy to use the proposed animated HM. However, the proposed animated HM still has some disadvantages. In particular, learners may have difficulties to grasp an initial overview of each topic at an early stage. It may be because the subject content has been divided into eight sections and only one section is displayed at a time. Therefore, they have to switch between each section and connect different topics by themselves if they need to get a completing understanding of the subject content.

**Collaborative learning**

Most of learners in collaborative learning think that using the traditional HM is efficient to switch between each topic so that they can easily locate relevant information. It may be because the traditional HM supports non-linear navigation with the backward/forward buttons provided by a browser. Such an advantage only exists when the map is compatible with the browser. As described in the previous section, the traditional HM, however, presents too much information so that learners may feel difficult to reach their learning goals. In addition, such a complex structure may lead learners to go to a wrong direction so that they have to spend much time finding relevant information for their tasks. Therefore, they cannot concentrate on their learning tasks.
On the other hand, most of them think that the proposed animated HM is well-structured so that they can explore the subject content step-by-step. A possible reason is that the subject content is divided into eight main sections in the proposed animated HM and the subsections are displayed above these eight main sections. However, learners may find it tedious to use the proposed animated HM. It may be because the proposed animated HM is not supported by the backward/forward buttons provided by a browser. Thus, they have to start from the first page every time when they need to go back to the previous section. This is also the reason why learners indicated the proposed animated HM is not as efficient as the traditional HM.

In summary, the results presented in this section indicate that our proposed animated HM does not have remarkable effects on individual learning while it can have greatly positive effects on collaborative learning. As described in the Theoretical Background Section, the difference between the individual learning and collaborative learning lies within the fact that only one student is involved in the former whereas more than one student are engaged in the latter. More specifically, discussion may take place in collaborative learning. Students may have diverse interests so that there is a need to have a tool to facilitate their discussion. The traditional HM presents massive information so students may find it difficult to share their interests. On the other hand, our proposed HM provides a well-structured categorization to help learners identify their interests. Furthermore, an arrow in the middle of the proposed animated HM works as a visual cue, which is useful to understand the interests of each other. In brief, our findings suggest that the proposed HM can enhance students’ communication in collaborative learning. Conversely, discussion does not take place in individual learning so the effects are limited.

Regardless of individual learning or collaborative learning, learners generally not only show more positive perception in using the proposed animated HM, but also demonstrate a better performance. These findings are in line with those of previous studies (Chen, 2002; Frias-Martinez et al., 2008), which highlighted that there was a close link between students’ learning perception and learning performance. More specifically, positive perception could enhance learners’ performance. These results reinforce the importance of a good design, which can not only enhance students’ learning performance, but also improve their learning perceptions.

Regardless of individual learning or collaborative learning, learners generally not only show more positive perception in using the proposed animated HM, but also demonstrate a better performance. These findings are in line with those of previous studies (Chen, 2002; Frias-Martinez et al., 2008), which highlighted that there was a close link between students’ learning perception and learning performance. More specifically, positive perception could enhance learners’ performance. These results reinforce the importance of a good design, which can not only enhance students’ learning performance, but also improve their learning perceptions.
Session 1
Usability of Interaction Design

Figure 8. Zooming in Hierarchical Map

Figure 9. The stretch text of the hierarchical map
Implications for the design of navigation tools

The results presented in this paper indicate that the proposed animated HM is generally appreciated by students. However, some weaknesses still exist so there is a need to improve the proposed animated HM. Firstly, students think that it is difficult to get an overview with the proposed animated HM at an early stage, regardless whether it individual learning or collaborative learning. One of the solutions to address this problem is to allow the learners to see an overall picture with the functionalities of zooming out and zooming in. More specifically, zooming out is to shrink each section of the map so that all of the sections can be displayed within the map (Figure 7). In contrast, zooming in is to magnify a section of the map so that the details of this section can be examined (Figure 8). In other words, the former is useful to show all topics for students who would like to get a global picture of the subject content. Conversely, the latter is helpful for students who would like to examine a single topic deeply. By doing so, zooming out can overcome the weaknesses of the proposed animated HM while zooming in can still keep the strengths of the originally proposed animated HM.

On the other hand, the results of our study showed that the traditional HM provided too much irrelevant information so students might experience cognitive overload. This is due to the fact that all of the topics are displayed within one map. The abovementioned zooming out may also help students address this problem, which can be solved by stretch text. More specifically, the map initially presents main sections only. When learners move the mouse to a particular main section where they are interested, a type of stretch text is used to present various subsections associated with this main section. Conversely, these subsections can also be hidden when the students intend to examine a single topic deeply (Figure 9). In other words, this approach increases the flexibility of the proposed animated HM so that learners can easily get a global picture of the subject content and examine a particular topic in details.

Conclusions

To address the problems of the traditional HM, this study proposes an animated HM. One of the advantages is that Adobe Flash was applied to create animations so that learners can switch between each section easily. The other advantage is that the proposed animated HM was carefully designed based on Nielsen’s five usability criteria, including “Efficient to use,” “Easy to learn,” “Few errors,” “Easy to remember” and “Pleasant to use.” An empirical study was conducted to investigate the effects of this proposed map on individual learning and collaborative learning. The effects of the proposed map on individual learning are not as obvious as collaborative learning. Nevertheless, learners positively reacted to the proposed animated HM in general. This suggests that incorporating usability criteria into the development of the animated HM can overcome shortcomings in the traditional HMs to satisfy users’ needs (Marsico & Levialdi, 2004).

Thus, this study contributes to the deep understanding of how to incorporate usability into the development of the HM. Furthermore, the design rationale can also be applied to develop other navigation tools, e.g., main menu. However, the proposed animated HM still has some weaknesses. Thus, several additional design approaches were proposed to address these weaknesses. In the future, the rationale of the design approaches can be used to improve the development of existing WBL systems and other Web-based applications, such as digital libraries, search engines, and electronic journals. Finally, it would be valuable to see whether learners’ performance and perception can be improved in such WBL systems and the Web-based applications.

Although this study shows fruitful results and proposes useful solutions for the improvement of the design of WBL systems and the Web-based applications, some limitations still exist. Firstly, this is a small-scale study. In the future, there is a need to conduct an empirical study with a large sample from a generic case. Furthermore, WBL systems are used by diverse learners. Therefore, further works should examine how individual differences affect learners’ preferences for the design solutions proposed in this study. The results from such further works can be incorporated into those of the present study so that a robust user model can be built for the development of personalized WBL systems.

Acknowledgements

The work presented in this paper is funded by National Science Council, Taiwan (NSC 98-2511-S-008-012-MY3, NSC 99-2511-S-008-003-MY2, NSC 101-2511-S-008 -010 -MY3).
References


Using Lean in the Flipped Classroom for At Risk Students

Shannon Flumerfelt and Greg Green

Oakland University, Rochester, MI USA // Clintondale High School, Clintondale Community Schools, Clinton Township, MI 48035-2633 USA // flumerfe@oakland.edu // greeng@clintondaleschools.net

(Submitted September 02, 2011; Revised December 10, 2011; Accepted February 03, 2012)

ABSTRACT

Schools are working to improve achievement through the examination of instructional practice and the use of instructional technology. This article provides informed commentary on the state of school reform and the need for continuous improvement, instructional improvement and instructional technology improvement. It also presents advocacy for the use of a continuous improvement system called lean as a toolkit for these improvement efforts. A discussion example of an at risk high school’s journey through continuous improvement and the use of a lean tool for analysis for improvement resulting in the innovative use of screen capture technology is shared to highlight one application of the lean framework presented.

Keywords

Lean, Flipped, School Improvement, Instructional Technology, Secondary Education

Introduction

The new millennium is well underway and demands are present for schools to meet the vision of relevance in the 21st century. For example, the Organisation for Economic Cooperation and Development (2008) describes the urgency for major changes to create consequence in schools. These include the need to emphasize lifelong learning and the learning process itself in order to prepare students for the knowledge economy, the need to personalize learning and use formative assessment as well as expanded information sources to prepare students for acceptable levels of global achievement, and the need for more professional collaboration networks to inform educators as to how to fully develop student potential. In addition, the Re-inventing Schools Coalition (n.d.) proposes a vision for relevance that includes self-directed learning and student empowerment, varied use of peer and teacher assessment, and student demonstrations of mastery as the basis for promotion. On a broader scale, Darling-Hammond (2010) advocates for policy reform that will alter schools to the point that “will enable students to learn how to learn, create, and invent the new world they are entering” (p. 3). Furthermore, the National Science Foundation (2008) calls for the use of cyberlearning to transform schools by providing students with “a mix of diverse content via the combined technological capabilities of the Internet, high performance computing, advanced networking, in-home electronics, and mobile communications” (p. 6). Certainly, these ideas are very exciting for the future of schools.

This article presents an informed commentary on the state of 21st century vision for schools. A broad overview of the need for school reform and the use of continuous improvement as a conceptual tool to drive instructional improvement is presented first. This is followed by a discussion of the potential of instructional technology as a tool for change. In addition, a brief presentation of a relatively new body of knowledge and practice for education known as lean is provided along with a description of how a demonstration site, an at risk high school, benefitted from lean analysis to highlight where to improve instruction through screencast technology (Green, 2011a).

The need for school reform and continuous improvement

Understanding what it means for schools to be consequentially significant in this century and further determining how to convert an educational system to such a state is a part of reform work that has occurred in the past two decades under programs such as Comprehensive School Reform (Borman, Hewes, Overman & Brown, 2003), the Coalition Campus Schools (Darling-Hammond, Ancess & Wichterle-Ort, 2002), and the Race to the Top Program (United States Department of Education, 2009). In addition, the National Education Technology Plan (United States Department of Education, 2010) is working to prompt needed change by asking education leaders “to enact revolutionary change; technology-based learning and assessment; engaging and powerful learning content; interactive, collaborative, visual and dynamic lessons…to be more relevant to life and work goals” (online). In this literature, there is advocacy to rapidly advance the bearing of instructional technology for the next century as
improvement work. While the need for school reform is widely recognized and practiced, the journey into the creation and sustaining of change is an area that schools struggle with even though there is shared concern in the educational community. The ability to understand how to implement and maintain change is hard. As Horsley and Kaser (1999) indicated, educational change has a low success rate for longevity and success. Saranson (1990) provided blunt commentary as to why this is so as he points out the self-sustaining nature of schools, based on long-established power systems and long-standing traditions as a countervailing force that, “…almost automatically rules out options for change” (p. 35). Furthermore, reform initiatives have fallen short of touted goals, producing confusion and disheartening key stakeholders. There are, therefore, gaps that are being recognized that exist between the hopes of school reform and the ability to engage in improvement (Dykema, 2002). Continuous improvement as best practice for change navigation in school is well established as a sustainable method of adding value to instruction (Schmoker, 1996). The concept of continuous improvement may well be the key to help schools bridge the divide between desired goals and current reality.

Continuous improvement requires a shared commitment to change. This dynamic is not difficult for schools to establish per se. But, where schools encounter difficulty is with the actual process of continuous improvement and how to do it. The use of continuous improvement is not easy and the issues of achieving school reform through continuous improvement are complicated with many variables and factors as Fullan (2001) described, “The big problems of the day are complex, rife with paradoxes and dilemmas” (p. 2). In an effort to better understand what educational theorists are recommending regarding continuous improvement, two specific factors of interest are examined next, instructional practice and instructional technology improvement.

**Instructional practice and continuous improvement**

In terms of instructional practice improvement, Elmore (2003) provided a description of what continuous improvement is like in a school, as a process of cultural alignment of “norms and values” that impact instructional practice. This description brings up an important concept regarding instructional practice improvement. That is, whether examined on a micro level, for one student, or on macro level, across a school, improvement hinges on a theoretical commitment to learning as a process of continuous improvement. The theories of organizational learning also help to identify the need for both personal/professional mastery as well as organizational attainment via the process of continuous improvement (Argyris & Schon, 1996; Senge, 1990).

Translating continuous improvement on an organizational level to impact instruction requires the willingness to critically examine reality against theoretical ideals. For example, an examination of instructional practice reveals potential misalignment with educational assessment best practice versus what is in use, for instance. Unfortunately, instructional practice norms and values tend to rely on the more efficient, exclusively cold cognitive model, which is conducive to high stakes testing. Ferrero (2005) described this shortfall of only measuring cognitive development only in this way as,

> We know, for example, that the mind constructs knowledge—that people learn by connecting new information to existing understandings and conceptual frameworks. We know that teaching needs to attend to both basic and higher-order skills, and to both cognitive and non-cognitive development. (p. 6)

So, while instructional reform is called for universally (White, 1997), it requires an organizational change process like Elmore (2003) provided, which allows for the realignment of norms and values that will improve instruction (Berry, 2011).

Therefore, in the case of any educational practice shortfall, it becomes evident that organizational change process theory is required for continuous improvement so that realignment can occur. For instance, the description that Schmoker (2006) provided depicts a clear picture of the organizational process of teachers implementing new instructional practice by using continuous improvement to better that practice and doing so under the norms of what “authentic literacy” (p. 51) learning entails with emphasis on both content mastery and high level analytical skill. The need to engage in continuous improvement regarding instructional practice is evident in regard to norms and values alignment around the process of learning itself, along with cognitive development and the increased use of
formative assessment. In other words, this work of aligning dispositions and instructional practice to theorists’ work defines what continuous improvement is.

Schools are interested in continuous improvement as both a cultural and structural solution to increasing student achievement. As Schmoker (2006) explained, continuous improvement is not an event, but a collaborative and transparent process of meeting regularly, piloting ideas, refining them, and finally publishing them while training people to implement them (DuFour & Eaker, 1998). Smylie (2010) reinforced this as well and outlines how highly regarded continuous improvement is as both organizational principle and practice.

Resnick (2010) indicated schools must be aware of instructional tools and resources at all levels (from district policy to classroom practice) along with good social capital management for instructional reform to occur. She clarified that this has not been done widely in schools and that there is a gap between the shared commitment to school improvement and the ability to implement. Resnick (2010) stated, “attempts to design education organizations and test those designs empirically in a continuous cycle of improvement are still rare” (p. 195). There is no formula for continuous improvement to be undertaken by school leaders. Elmore (2006) explained:

One does not ‘control’ school improvement processes so much as one guides them and provides direction for them, since most of the knowledge required for improvement must inevitably reside in the people who deliver instruction not in the people who manage them. (p. 58)

Continuous improvement as leadership practice is not unspecified, however. The work of Marzano, Waters, and McNulty (2005) highlighted as principle-based findings which leadership behaviors are helpful to second order change agency and which are not. Again, though, they are careful to explain that they are not presenting a prescriptive approach to continuous improvement or second order change. Blankstein (2010) endorsed this conceptualization by describing continuous improvement as careful, thoughtful and organic work fostered by collective reflection, collaborative work and leadership development.

The school example described later will focus on how continuous improvement has changed the traditional use of time on task for instruction and created new opportunities for focusing on the process of learning and summative assessment work. This was operationalized through the use of screencast technology as instructional technology improvement. Before the example is presented, the need for continuous improvement in instructional technology is described next.

**Instructional technology and continuous improvement**

Central to the vision of the 21st century school is the use of instructional technology. As Daggett and McNulty (2005) stated, “More extensive scientific and technological advances will occur in the next few years than have happened in the last two centuries. Dealing with these advances requires a different education system from the one in which we were educated.” (p. 12). Others envision instructional technology use in ways that provide multiple learning options for students, such as Collins and Halverson (2009) described as, “the new seeds of an education system forming in the rapid growth of learning alternatives, such as home schooling, learning centers, workplace learning, and distance education.” (p. 3). This prediction, if actualized, will surely force an improvement response from the current education system. Furthermore, the National Education Technology Plan (U.S. Department of Education, 2010) set out a vision to use instructional technology as a tool-enabled approach to reform education.

Unfortunately, to date, instructional technology implementation in schools has a scattered record. Zhao and Frank (2003) described the situation as confounding.

Concerns about the slow adoption of technology by teachers are not new. The phenomenon has been studied from different angles from case studies (Cuban, 2001; Shofield, 1995; Zhao et al., 2002) to historical analysis (Cuban, 1996), to large surveys (Becker, 2000a, 2001) (p. 808).

Zhao and Frank (2003) analyzed why this has occurred and indicate how to correct this shortfall by concluding that schools should be viewed as systems and that instructional technology implementation should be properly placed taxonomically within that system as a source of organic organizational growth or continuous improvement. This conceptualization of technology implementation is contrasted with the notion that technology use can be mandated,
funded or supported as an external and disruptive force to an existing system. In other words, their research posited that technology implementation is most effective when implemented as continuous improvement.

To further understand what the 21st century school looks like in regard to instructional technology, Morris and Hiebert (2011) described the need for “shared, changeable knowledge products” that solve common instructional problems by gradual improvement from multiple and accessible sources. The National Science Foundation (2008) encouraged the development of instructional technology under the broader category of cyberlearning, “Cyberlearning has tremendous potential right now because we have powerful new technologies, increased understanding of learning and instruction, and widespread demand for solutions to educational problems” (p. 5). The ability of schools to empower students to navigate learning pathways, make instructional choices, and receive assessment feedback; to enable teachers to facilitate that process of learning with student data management systems and differentiated instruction; and to provide schools with collaboration networks under enterprise architecture is currently under development (Hillberg, Flumerfelt, VanTil & Tierney, 2011). The National Science Foundation (2008) confirmed the importance of moving to whole scale reform in regard to instructional technology as, “Learning support systems can and will be organized along very different schemes than they are today, given the computational services made possible with cyberinfrastructure advances.” (p. 37). It is evident that the future of schools is hopeful in regard to the use of instructional technology as continuous improvement occurs in this area.

The next section provides a brief introduction to lean, a philosophy and method useful when engaging in continuous improvement. A lean tool, the value stream map, is demonstrated as a visual management tool for navigating continuous improvement for the demonstration site presented later that utilized screen capture technology.

The application of lean for continuous improvement in schools

The body of knowledge and practice known as lean is widely regarded in various sectors as best practice for continuous improvement (Robson, 1991; Slater, 2007). Because of this, lean is gaining interest in the educational sector as a useful organizational philosophy and administrative toolkit bundle (Stecher & Kirby, 2004; Barney & Kirby, 2004; Zivkosky & Zivkosky, 2007; Magua, 200; McMahon, 2006; Pawley Lean Institute Website, 2011a; Balzer, 2010). Lean is defined as an approach that requires the commitment of the technical, social and human capital of an organization to continuous improvement for the purpose of identifying distinct ways to create value as determined by the customer and to eliminate waste based on thoughtful examination of its root causes (Womack, Jones & Roos, 1990). Bhasin and Burcher (2005) were clear that to be successful, lean must first be philosophically and culturally framed and then operationalized. They stated, “While lean is concerned about reducing waste, it is also about changing corporate culture” (p. 58) and specify continuous improvement as one of twelve organizational practices for lean manufacturing.

A landmark study conducted in 1988 with a $5 million grant examined best practice from the international automotive factory floor and named that body of knowledge lean (Womack, Jones & Roos, 1991). With a long history of development, originating with the production of goods and lasting over centuries, the term lean is fairly new, but it represents centuries of thinking and tools that advance performance, referred to by Womack (Lean Enterprise Institute, 2010) as “waves of improvement” (online video). Lean is an organizational philosophy and operating system heavily embedded in the total quality work of W. Edwards Deming (Dennis, 2006) and the continuous improvement cycle of Shewhart (1980), the Plan-Do-Check-Act cycle which Deming popularized. Further, Emiliani (2008) emphasized that lean uses two weighted principals; the lighter one being continuous improvement and the weightier one being respect for people. Ransom (2007) in a Lean Enterprise Institute resource testified that lean culture and continuous improvement in turn produces desirable bottom line results as well, “Essentially, it [lean implementation] is a virtuous circle. I would stipulate that the adoption of a lean culture will improve financial performance.” (Slide 2). The most sophisticated lean enterprises have embraced the practice of enacting lean so that it moves from continuous improvement of internal operations to the full stakeholder chain (Burton & Boeder, 2003). In all, lean results in three aims, stakeholder engagement in continuous improvement, stakeholder application of improved paradigms and processes, and organizational respect for empowered stakeholders in the continuous improvement process (Liker & Hoseus, 2008).

To understand the application of lean to assist continuous improvement efforts, there needs to be recognition of the differences between businesses and schools between the structural and resource goals in private versus public
organizations. Furthermore, lean application requires careful translation through contextualization of best practice from one sector to another. Collins’ (2005) view is helpful in this regard as he stated, “In business, money is both an input (a resource for achieving greatness) and an output (a measure of greatness). In the social sectors, money is only an input, and not a measure of greatness” (p. 5). Lean is a business concept, initially used to drive total quality management in production (Womack, Jones, Roos, 1991). It has been and continues to be translated with for K-20 education under initiatives such as Lean Thinking for Schools™ (Pawley Lean Institute, 2011b) and Lean Engineering (Kahlen, Flumerfelt, Siriban-Manalang & Alves, 2011) among others. To date, lean currently is understood as an emerging, yet helpful philosophy and strategy for improvement in education (Lean Education Academic Network, 2011; Flumerfelt & Banachowski, 2011; Flumerfelt, 2011).

An example of lean, continuous improvement and instructional technology

In terms of specific uses of lean in schools to assist with continuous improvement, problems such as ineffective remediation, the lack of developmentally appropriate learning opportunities for students, and inadequate funding provide excellent opportunities for exploring lean thinking and applications. Consider, for example, the concerns expressed over standardized testing and what this means for culturally diverse populations. In terms of lean thinking, the basis for a successful lean enterprise is, in fact, founded on the utmost respect for all stakeholders, especially the employees and the customers, per Taiichi Ohno, the architect of the Toyota Production System, as described by Liker and Hoseus (2008). This means that lean organizations devote time and resources to understanding what is of value to those various stakeholders and then improving their systems to better meet those needs. For culturally diverse students, especially those at risk, our response in education has largely been of the mindset of remediation of students—“let’s fix these students!” Lean thinking is in opposition to remediation processes for students and supports in its place improvement of instructional processes. What is proposed in lean approaches is to fix the root cause of the problem in processes, rather than “fixing” people. Lean problem solving focuses on engaging people to identify process improvements and to create those solutions as a “Let’s fix the processes that do not work for these students!” approach. Lean does not use a one-way standardized formula derived from the top-down, but rather in-house solutions. So then, lean could be extremely helpful for students who suffer at the hands of the shortcomings of our current educational system. The tools of lean allow stakeholders to examine the educational system closely, develop collaborative solutions and participate in continuous improvement processes (Villareal, 2011; Brown, 2011). For the inequities in education that plague some of our students, lean is a viable process improvement approach to be considered.

Lean can be used by schools to examine processes for improvement ranging from the core technology of delivery of instruction to administrative support. For instance, value stream mapping (Keyte & Locher, 2004) is a lean tool that can be used to solicit the views of key stakeholders, such as students, teachers, parents, policy makers, administrators and boards, in regard to what is of value in the instructional delivery process. A student’s instructional day can be mapped out looking at allocations of time and resources for various activities. Based on the views of key stakeholders, decisions would be made as to what is of value during that instructional day and what is not. What is of value is kept and what is not is either improved so that it becomes valuable or it is eliminated.

![Figure 1. Key elements for lean value stream map current state, kaizen, future state](image-url)

360
Since value stream mapping is one lean tool that might be useful to facilitate continuous improvement of instructional delivery, an example is provided in Figures 1-4. Value stream mapping is illustrated a three step process of (1) visually mapping (Figure 1) an instructional process in its current state (Figure 2) through the eyes of the stakeholder, the student; (2) examining the map for ways to increase value to the student (known as kaizen) and indicating those on the current state map (Figure 3); (3) and finally improving the process from kaizen to create a future state map (Figure 4). Value stream maps examine each step in a process by assigning valued metrics throughout and identifying where improvements can be made. These maps are typically drawn right to left and use common icons, noted in Figure 1.

Figures 1-4 demonstrate how value stream mapping is used to understand what a high school with a growing at risk student population can do to improve instruction (TechSmith, 2011). The example in this article is not presented as a case study, but, rather, as a discussion example of what one high school did to improve instructional practice using continuous improvement and instructional technology. Figure 2 illustrates traditional instructional practice in one high school class (55 minutes) delivered to the at-risk student and how each activity in the class fares against five metrics, important to the at-risk student. These five metrics are (1) the ratio of time allocated to task/relational activities, (2) the ratio of time allocated to passive/active learning, (3) the amount of time allocated to new learning opportunity, (4) the amount of time allocated to individualization, and (5) the amount of time allocated to...
differentiation. For each class activity or process box, a metric box indicates the status of these five metrics. For each metric box, the results can be totaled and then compared to the ideal results in the key provided. If the actual results fall short of ideal results, then possible points of improvement are identified, planned for, deployed, and reassessed, enacting continuous improvement.

Continuous improvement occurred at the demonstration site, the at-risk high school, around several points, as potential areas of improvement were highlighted by kaizen bursts (Figure 3). As each individual process step was examined, improvement options stemmed from the desire to deliver more value and based on achieving better metrics for the student for each step in the process. As indicated, the school decided to use instructional technology to “flip” (Bergman & Sams, 2011) the use of classroom time away from lecture to facilitation of individualized and differentiated instruction. This greatly impacted the metrics by enabling teachers to facilitate learning more during class time and to deliver content outside of class time. Furthermore, the use of learning management software to facilitate online facilitation and support was also identified, resulting again in improvement in classroom dynamics that lead to better metrics.

As the process improvement in the future state, Figure 4 depicts how the school created value by using digital screen capture technology and online learning techniques. Screen capture is a software tool that enables a teacher to record
anything that is on the computer screen, annotate it and add voice over. Once the recording was processed, it was uploaded directly to a video sharing site for distribution through the internet. Online learning was used to provide resources, learning groups, student monitoring and instructional facilitation. The metrics in Figure 4 indicate the impact of these improvements.

![Diagram](image)

Figure 4. Sample classroom instructional delivery future state value stream map

As the value stream maps in Figures 2-4 demonstrate, this school addressed the gaps in traditional high school instructional practice by enabling teachers in the classroom to “flip” lecture time into time for higher level engagement by students with teachers. This improvement allowed students to increase processing time substantially and to do so in an environment in the classroom, where teacher and peer support was available. The advantages to digital screen capture also allowed students to review, as needed for remediation or intervention, or not, if mastery was evident. In standardizing the delivery of content through instructional technology, the ability to individualize and differentiate instruction for the student as an instructional practice improvement was evident. This change also created more instructional time overall by providing students with a task they enjoyed and could control, deciding when and where to listen to the daily lectures.

The metrics in Figure 4 illustrated these improvements. This depiction highlighted that continuous improvement work using the lean tool, the value stream map, used metrics of importance to the stakeholder. Of course, these
metrics may not create the only benefits for the improvement solution. For example, this instructional technology solution also provided teachers with the flexibility to create learning groups based on student needs and it allowed the school to select lecture content creation based on strength of delivery by teacher expertise. A teacher-based value stream map could be used to highlight these ancillary improvements.

The initial results of this flip improvement are encouraging. The school worked from a small pilot of 23 at risk second semester students in a government class and used a control group for comparison. In the experimental flipped class, the students increased their online engagement and homework rates from 75% to 100%. Students’ successes increased by 11% in the flipped class. This resulted in an elimination of all students’ class failures. Following the pilot, this improvement has been implemented with the 9th grade class overall and year over year improvements are noted, such as discipline events decreasing by 66%, while failure rates reducing in mathematics by 31%, English by 33%, science by 22%, and social studies by 19%. This demonstration site of continuous improvement in instruction using instructional technology (Green, 2011b) does provoke the need for further study, as these results are impressive.

**Summary**

Schools are striving to continuously challenge current instructional practices in order to produce improvement, not just change for change’s sake, but by engaging in value added improvement. The ability to engage in continuous improvement is a hallmark behavior for schools striving to achieve the vision of the 21st century school. Daggett and McNulty (2005) endorse these approaches to school reform as a continuous challenge to the status quo as follows,

> As the demands to raise standards have become steeper, schools tend to rely on tried-and-true curriculum content and teaching approaches. However, this old methodology was intended for an education system whose mission was to select and sort students, not to move all students to high levels of proficiency. (p. 13)

In other words, as familiar and traditional models of instructional practice come under the lens of reform, schools must effectively engage in continuous improvement. This informed commentary has attempted to link extant literature regarding the need for continuous improvement, instructional improvement and instructional technology improvement as a theoretical basis for creating interest in the translational work of the use of lean as an improvement tool in the educational sector.

Further, an example of one high school’s journey into continuous improvement and how value stream mapping was used to work through process improvement was presented. The academic achievement and behavioral referral improvements experienced in a pilot of flipped classroom instruction using screencast video technology were impressive. It is hoped that this modest example will prompt further scholarly evidence-based practice and research to inform the field regarding viable organizational approaches to school improvement that link theory with practice.

**References**


Using E-readers and Internet Resources to Support Comprehension

Sandra Wright, April Fugett1 and Francine Caputa
University of Tulsa, Tulsa, OK74104 USA // 1Marshall University, Huntington, WV, USA // sandra-wright@utulsa.edu // fugett5@marshall.edu // fmcaputa@gmail.com

(Submitted September 02, 2011; Revised December 09, 2011; Accepted December 30, 2011)

ABSTRACT
The advancements of technology have led to the use of electronic reading systems for digital text. Research indicates similarities and differences in reading performance and comprehension in digital formats compared to paper formats. This study compared vocabulary understanding and reading comprehension scores from two reading sources (electronic story book and paper-based book). This study also evaluated the use of reading resources available (dictionary, thesaurus, word pronunciation) between the two reading methods. An AB experimental design consisting of three females currently enrolled in the second grade, between the ages 7 years 0 months to 8 years 11 months without an identified disability, was conducted between two reading methods (paper versus electronic book source) with the participants serving as their own control in both conditions. The results of this study conclude that although vocabulary and reading comprehension is consistent between the two reading methods, students are more likely to utilize reading resources when engaged with digital text. This article supports that comprehension of written materials remains unchanged for students regardless of presentation method (print versus digital). It also provides evidence that supports students who are typically developing demonstrate more willing to utilize reading resources when technological advancements are integrated into reading practices. Further research is needed in order to determine if this trend is consistent for children with a diagnosed receptive language or reading disorder.

Keywords
Literacy, Language, Technology, Digital Text, Reading Comprehension, Vocabulary

Introduction

The National Center for Education Statistics (NCES) showed that in 2003, 40% of fourth grade students in the United States scored below the basic level of reading expected for their age (Ogle, Sen, Pahlke, Jocelyn, Kastberg, Roey, & Williams, 2003). The impact of increasingly available electronic devices, and their use within society, could be a contributing factor to the decline of voluntary reading time for children. Children of the new generation are being referred to as digital natives (Prensky, 2006). Digital natives are born into an electronic world in which it seems their brains are immediately wired to acknowledge electronically operated devices, such as television remote controls, computers, and music players. Researchers claim that activities for children of this generation are more electronically driven than the activities of any other generation, and young children today are more intrigued by electronic devices than any other type of entertainment (Kinzer, 2003; Larson, 2009; Larson, 2008; Prensky, 2006). Digitally operated devices are thought to capture the interest of the American mind more than paper-based activities (Bennett, Maton, & Kervin, 2008; Prensky, 2006; NEA, 2007). This transition is evidenced through the availability of electronic books (e-books). By altering the mode of reading material from traditional paper-based reading to electronic book (e-book) reading, interest in reading may increase, and thus be an effective solution to promote literacy activities both in and out of educational settings. Numerous studies illustrate the benefits of e-book reading as an effective means to improve the literacy skills of children, especially reading comprehension when assessed using multiple choice questions (Korat, 2009; de Jong & Bus 2002; Grimshaw, Dungworth, McKnight, & Morris, 2007).

Literacy is the key to success in a literate culture and an essential element to the educational development of students (Hay & Fielding-Barnsley, 2009; Moats, 2000; NEA, 2007; Strickland & Morrow, 1989). Well-developed reading and writing skills are associated with higher levels of academic achievement. Before a student can begin to develop literacy skills, several language structures must be developed in listening and speaking skills (McLaughlin, 2006). The development of literacy skills must exceed the production of pronouncing a printed word, or constructing letter sequences through writing, there must be a level of comprehension of the language embedded within these formats. This is why the development of grammar, semantics, metalinguistics, and pragmatics should be apparent in spoken language production and auditory understanding prior to the initiation of reading instruction. Reading and writing are complex processes that demand appropriate instruction for satisfactory acquisition of the skills necessary to
becoming a competent reader (Gleason, 2005; Moats, 2000; Strickland & Morrow, 1989). Learning to read is a complex process that is not easy to teach and consists of several components. The ability to read, and understand what is read, is directly impacted by a person’s expressive and receptive language abilities. Spoken and written language skills are the ways in which someone expresses him- or herself, whereas listening and reading are means of receiving language. Developmentally, children first make gains in their listening skills that help generate new spoken vocabulary. Children must then take their spoken language knowledge and translate it into literacy development, specifically reading decoding and reading comprehension skills. The recognition of a word, and knowing that it has a function, stimulates the ability to understand the meaning of multiple words in the context of a sentence, which is called semantic knowledge (Gleason, 2005). At early literacy levels, a person is able to use prior knowledge to decode words, understand the parts of the word and how the word relates to surrounding words within the given context, and connect a meaning to the word based on its’ context. Semantic knowledge opens the door to reading comprehension and interpretation of text. With the ability to connect meaning to print, a child is able to read stories more fluently and eventually read for educational purposes and learning new knowledge (Gleason, 2005). Direct instruction for reading usually starts during kindergarten and first grade; however, literacy learning is a lifelong process that begins indirectly during infancy and naturally continues throughout adulthood (Gleason, 2005). Prior to school enrollment, children’s literacy skills begin to build through interactions with their environment. At home and in community settings, most children are surrounded by print material such as alphabet blocks, toys, newspapers, magazines, grocery lists, public signs, logos, and labels (Strickland & Morrow, 1989). Exposure to print, active engagement with literacy-rich activities, and facilitation by a caregiver are important experiences that prepare one for literacy development.

The five essential components of reading include phonemic awareness, phonics, comprehension, vocabulary, and fluency (Gleason, 2005; Hay & Fielding-Barnsley, 2009; Moats, 2000). When students in upper elementary and middle school grades demonstrate reading comprehension difficulties, the area of difficulty is typically weak phonological processing skills and the inability to decode words; skills they failed to master in their early years of literacy development (Moats, 2000). As a result, these students frequently skip over unrecognizable words, lack the ability to read fluently, and, therefore, struggle to comprehend age appropriate stories and educational material (Moats, 2000). Efficient reading skills do not develop easily for all children, and, within the school system, students often fail to receive special reading intervention when it is most needed. Children who experience numerous difficulties in reading may begin to associate negative feelings with reading activities. Children will frequently attribute their reading difficulties with low ability that can result in reduced self-esteem and therefore a lack of motivation to read (Kamhi & Catts, 1991). As a consequence of such deficits, a person who lacks motivation to engage in reading academic material activities may increasingly continue to fall behind in literacy skills and, thus, have difficulty advancing with higher educational knowledge. Reading intervention must be addressed aggressively. All children do not respond to repeated failure in the same manner, therefore alternate approaches to the teaching of reading skills should be utilized (Kamhi & Catts, 1991). Teachers and speech language pathologists need to focus on phonological skills, age-appropriate vocabulary, and comprehension strategies, while presenting instruction in an appealing manner that motivates the child to read (Gleason, 2005; Moats, 2000).

Technology is argued as a solution that will close the gap between teachers, and the traditional method of teaching, and students of the new generation. E-books provide an opportunity for teachers and children to connect and share knowledge, build relationships through literacy developing activities, and enhance stimulation in literacy enriched environments (Strickland & Morrow, 1989). Larson (2008) argues that altering traditional reading workshops to electronic reading workshops is crucial, as well as, educating teachers on how to implement technology and effectively use e-books within an elementary classroom. If e-reading workshops are not available for an educator, public librarians are a proficient resource for teaching adults, educators, and children digital literacy skills and showing educators how to best take advantage of literacy-learning options offered by e-books (Borawski, 2009). There is also concern regarding the distractibility of electronic sources used for e-books that may distract the reader’s attention away from the text (de Jong & Bus, 2002). There are some concerns regarding the use of digital text, such as decreased reading speed, inability to accurately scan or notice details when screen reading, and distractibility from other available activities; however the benefit of readily available reading resources outweighs these concerns. Moreover, numerous studies illustrate the benefits of e-book reading as an effective means to improving the literacy skills of children, especially reading comprehension (Korat, 2009; de Jong & Bus, 2002; Grimshaw et al., 2007). Many teachers and researchers agree that e-books aid in the comprehension of reading and enhance literacy acquisition for typically developing children, as well as, children with learning disabilities of communication disorders (Black, 2010; de Jong & Bus, 2004; Korat, 2009; Shamir, 2009). Teachers help facilitate the development
of cognitive processes necessary for reading comprehension skills, especially for early readers. These cognitive processes are also developed through a variety of adult-child interactions. Unfortunately, some children lack, or have significantly reduced, adult-child literacy interactions in their home and school environments, thus inhibiting the acquisition of age-appropriate reading skills. Reading difficulties in children deprived of such reading interactions is frequently due to word-decoding failure, which can lead to poor vocabulary skills and thus difficulty with comprehension of higher-level reading materials. Adult-child interactions allow the child an opportunity to have instant oral-text narration, word-pronunciation, definitions and other useful literary information provided that are similarly available in e-book reading systems. These resources, as provided by electronic readers (e-readers), benefit emergent readers when parents or teachers are not readily available, as well as, late-developing older readers who desire privacy that prevent them from requesting help (Grimshaw et al., 2007). For struggling readers, and typically developing readers alike, e-book reading can provide a means for increased interest in literacy activities (Grimshaw et al., 2007).

Grimshaw and colleagues (2007) conducted a study that investigated reading comprehension and interest of reading of students based on the source of reading material (electronic versus paper-based). The research team analyzed two data sets. The first compared comprehension scores of two groups of students reading the same story through different sources. One group of students read the story digitally, while the other read a traditional paper-based book. Results illustrated no significant difference in overall reading comprehension abilities based on reading source. The second data set compared the comprehension scores and enjoyment level of three groups of students after reading the same reading passage either electronically without narration, electronically with oral-narration, or in a paper-based version. Results of a one-way analysis of variance (ANOVA) demonstrated that comprehension skills are significantly higher for electronic book (e-book) readers when the reader utilizes the available e-book resources, such as oral-narration. Based on student report using a 3-point Likert scale, the enjoyment level of reading was higher for the students reading the e-book with oral-narration versus those reading e-books without narration or those reading the printed version. The research team conducted their study with each participant individually, controlling for his or her reading level, book interest, age, gender, novelty of texts presented, and time given to complete the reading (Grimshaw et al., 2007). These results support that students’ comprehension of text is unchanged when reading digital text compared to print. Although the time needed to complete the reading of digital text increased compared to print, students indicated a higher level of enjoyment while reading in this format.

De Jong and Bus (2002) compared children’s attention to text based on the method of story narration (adult, e-book with games, e-book without games, audiotape) for 48 participants matched on age, socioeconomic status, reading level and absence of disability. The researchers provided an equal amount of time for reading in each of the conditions for the participants. The participants completed tasks related to the study individually, with only one examiner present to ensure that only the condition targeted was available to the participant (de Jong & Bus, 2002). Results from this study revealed that features offered through e-books distract the reader’s attention away from the text, especially when games are not restricted on the e-book reader, but overall ability to recall the story was similar regardless of delivery method. However, Korat (2009) conducted a study that demonstrated significant benefits of e-book reading when word meaning and comprehension skills were evaluated using yes or no questions and compared to print-based book reading. Moreover, in another comparable study by Korat and Shamir (2007) results concluded that children receive similar literacy benefits when they read independently using an e-book reader and when an adult reads to them using a book in print. It is important to note that in both studies, Korat (2009) and Korat and Shamir (2007), children using the e-book were limited to the available resources on the e-book reader by omitting access to gaming programs or application options. Participants involved in these studies were all within the emergent literacy age-range; an age in which it is normal to receive outside support to interpret and comprehend text. These participants were asked to engage in the reading tasks, controlled for age of participant, reading ability and interest level of the participants, individually but while being in groups of three. These studies indicate the potential of digital text to enhance literacy development, especially those in kindergarten and first grade (Korat, 2009; Korat & Shamir, 2007).

The American Speech Language Hearing Association (ASHA) (2010) stated that an educational reform addressing literacy achievement within the schools is needed in response to the decline of academic scores and school drop-out rates over the past decade. To improve the overall academic success of America’s current students, educators are encouraged to place direct attention on literacy achievement (ASHA, 2010). The demand to increase overall academic success within the schools is addressed through the emphasis of literacy acquisition activities, thus the role
of literacy intervention for speech language pathologists is escalating. Therefore, discovering an alternative means to
story reading that enhances children’s reading interest and comprehension skills is important to this field of study.

In order to meet the needs of literacy development for today’s children, several researchers and educational
authorities argue that traditional classroom teaching styles must be modified (Black, 2010; Kinzer, 2003; Larson,
2009; Larson, 2008; Prensky, 2006). Schools need to act on this notion and, instead of continuing traditional means
of teaching, need to utilize technology within the classroom as a foundation to reinforce the interest and motivation
of literacy engagement. Several teachers and researchers argue that by adopting technology into the classroom,
educators will provide more efficient and relevant education to their students (Bennett et al., 2008; Kinzer, 2003;
Larson, 2009; Larson, 2008; Prensky, 2006). Teachers are increasingly encouraged to use various forms of
technology within the classroom which the newly revised NETS [National Educational Technology Standards] for
teachers provide a framework for transitioning schools from industrial to digital places of learning (Larson, 2008).
Such technologies include PowerPoint presentations, online communication systems, online research, electronic
discussion boards, and multimedia teaching techniques within the classroom (Kinzer, 2003). Teachers are
encouraged to download e-books, use digitally interactive reading activities, and learn to how to best utilize these
sources for literacy development (Borawski, 2009; Larson, 2008).

E-books are available through the Internet or CD-ROMs. E-books available through the Internet are usually
purchased through online bookstores or borrowed free from electronic libraries or other public resources (de Jong &
Bus, 2002). CD-ROM books are purchased through local bookstores, toy stores, drug stores, catalogs, computer
stores, or borrowed free from public resources. The system used for reading, a computer, tablet or palm-based device
(e.g., Amazon Kindle, iPad, Sony’s Reader Digital Book, or PDAs), determines the resource for locating electronic
texts: Internet or CD-ROM. Most e-books include multimodal tools such as written text or oral narration, music,
sound effects, dictionary, thesaurus, hotspots, and animation (Korat, 2009). Most e-books additionally have useful
educational tools for highlighting, bookmarking, note making, search features and other valuable features for more
advanced, educational reading (de Jong & Bus, 2002).

Although most schools currently do not have the financial means to purchase PDAs, tablets or other hand-held
electronics, most schools offer computers that teachers may use for classroom lessons. Schools may either have a
room or hallway designated for computers, or may only have computers within the school’s library. Schools usually
have access to enough computers for each student within a single class, or to share with a partner. Teachers can have
students access e-books through an e-book website, or the educator could download a CD-ROM story onto the
computer prior to the classroom session. Larson (2008) reported that it is the responsibility of educators to integrate
the use of technology into literacy instruction and the entire language development curricula.

Purpose of the study

The purpose of this study was to determine whether children utilize readily available technology features available
on e-book readers (i.e., animation, highlighting, dictionary, thesaurus, etc.) to support overall reading
comprehension. This study did not intend to validate that e-books should replace paper-print books for children but
simply to discover if e-books provide an alternative option to reading which may be more effective than the
traditional method. This study intended to replicate the findings from previous research in regard to reading
comprehension similarities regardless of reading format, but support that students may have greater independent
learning through the use of embedded reading reference materials. For the purpose of this study, attention focused on
the multimodal options offered by e-books that support early developing reading skills: thesaurus, dictionary, and
text-to-speech capabilities. Such reading resources support comprehension of texts that are otherwise not easily
accessible to children when reading traditional paper-based books. The interest and enjoyment in reading during the
two reading methods (electronic versus paper) was also explored. It is hypothesized that with the integration of
technology in reading instruction, participants will increase their overall reading comprehension.

Methodology

Participants were recruited through word-of-mouth through the Mary K. Chapman Center for Communicative
Disorders on the University of Tulsa college campus. Eligibility was determined through completion of a parent
questionnaire (See Appendix A), hearing screening, language assessment, and reading assessment. Students that
qualified based on the parent questionnaire were assessed using a reading screener (Slosson Oral Reading Test-Revised (SORT-R)), and standardized language assessment (Clinical Evaluation of Language Fundamentals-Revised (CELF-R)) at The University of Tulsa communicative disorder clinic by the researcher under the supervision of a certified and licensed speech-language pathologist. The participants were the first three students who met all eligibility criteria and agreed to participation in the study through parental informed consent and assent procedures.

Participants

This study consisted of three females currently enrolled in the second grade, between the ages 7 years 0 months to 8 years 11 months. All participants scored within normal limits for their age in language skills according the CELF-R (Semel, Wiig, & Secord, 1995), and within normal limits for their age in reading skills according to the SORT-R (Slosson, 2003), as assessed by the researcher. Parent reports indicated that all participants had hearing and vision abilities within normal limit, and did not have any history of special education services or recommendations for special education services. Participants were all from middle socioeconomic status families and English was indicated as the primary spoken language in the home. All participants had access to computers and video gaming systems at home during participation in this study. According to parent report, none of the participants had previous exposure to the specific stories used for this study or history of using any type of e-reading system except for e-mail or Internet browsing. In order to insure protection of the participants this research was approved by the institutional review board at the University of Tulsa and employed both informed consent by the parents and assent of the participants. There were no expected risks to the participants in this study. Interestingly, participant’s reported that their reading experience was enjoyable, particularly when reading electronically. A small sample size was collected in order to make a book club more feasible, and would be more prevalent in an applied setting.

Research design

An AB experimental design was conducted between two reading methods with the participants serving as their own control in both conditions. Each participant engaged in reading sessions that met four times over a three-week period for one and a half to two hours per session. During the reading sessions lead by the researcher at The University of Tulsa communicative disorder clinic, the participants read one story using a traditional paper-based book and one story using an electronic-based book. The stories were equally divided across the four reading sessions; the first and second reading session focused on the paper-based book and the third and fourth session focused on the electronic book. Each story was divided into four segments and the participants were required to read each segment during a time period. Prior to the first and third session, participants were given instructions on the use, and recording process of the available resources (dictionary, thesaurus, and word pronunciation) for each reading method. Participants read simultaneously, in the same room which was set up similar to a living room with chairs, tables, and couches, and in individually chosen “reading-spots.” Variables that were considered to be influential in affecting outcomes were alertness level of participants, interaction of participants, and amount of time allotted for reading. The sessions were designed to occur on days in which the participants did not have to attend school in order to limit the amount of general fatigue from a long school day. The participants were not allowed to engage or interact with one another throughout the timed 15-minute silent reading period, or while completing the comprehension quizzes. The researcher designed “reading-spots” with props such as a beanbag chair and a blanket sprawled on the floor in order to give each participant his or her own space, and remove the possibility of interactions. The researcher remained in the room to monitor the participants throughout the independent reading time to ensure there were no interactions among them during the reading time or that the participants did not refer to the text while completing their comprehension quizzes. Participants independently completed comprehension quizzes in individually assigned areas within the same room. A description of each reading session is found in Appendix B. This method is similar to other studies whose aim was to assess text comprehension (Cross & Paris, 1988; Ray & Beldon, 2007; Senechal, 1997).

Instruments

Storybooks

A book via an e-reader, specifically the Apple iPad, and a book via paper-based print were used during this study. One story was chosen for each reading method. Both stories were matched equally in level of reading difficulty as
determined by Scholastic, reading interest based on gender and grade of participants, genre (fiction), and page length. According to scholastic’s reading level system, both stories qualified at the 2.3 reading difficulty level, pre-k to third grade reading interest level, categorized under the genre of mystery, and met the page-length between 50-65 pages.

**Comprehension quizzes**

A total of four reading comprehension quizzes were given to each participant during the study; two quizzes based off the paper-based story and two quizzes based off the e-reader story. The researcher constructed the quizzes. Each quiz was equally matched in style and type of questions asked, and consisted of four multiple-choice questions; two term-defining questions, one implicit question, and one explicit question (See Appendix C).

**Resources available for use**

A paper-based dictionary, a thesaurus, and the option to ask the researcher questions were provided while the participants read the paper-based story. While reading via electronic source, participants had access to use the dictionary, thesaurus, and word pronunciation tools offered on the iPad. Prior to beginning the story on e-readers, all participants were instructed on how to utilize the technological resources for reading.

**Data analyses**

Data analyses consisted of collecting the number of times literacy resources (dictionary, thesaurus, and word pronunciation) were used while reading for each of the two reading methods, and assessment of overall comprehension (based on frequency of correct responses) of the read materials for each reading period. A self-reporting tally system was used to record the total use of reading resources by each participant. Video recordings were used for viewing after intervention as a reliability check by the researcher and a trained outside observer. A comprehension quiz was given to each participant after each reading session to measure reading comprehension. A $\chi^2$ McNemar Test for significance of change is used for dependent sample designs and was calculated for each of the two hypotheses because of the small sample size, the AB design, and the collection of frequency data (Hinkle, Wiersma, & Jurs, 2003). Further, the print source was treated as a baseline of comparison to the electronic texts thus simulating a pretest-posttest design. It was felt that the $\chi^2$ McNemar Test for significance of was the best fit for statistical analysis.

**Results**

Two sets of data were analyzed. The first data set was the number of times literacy resources were used by type of reading method. The second set of data was the overall reading comprehension of text-based quizzes by reading method. Resource use data was calculated using a tally system. Analyses determined how frequently the participants used available literacy resources per method of reading. The frequency of use between the two conditions, paper-based and electronic reading was then compared. As indicated by Figure 1, the participants utilized more literacy resources when reading in an electronic format.

Data regarding participant’s reading comprehension was collected and analyzed using a point system with two, four-point quizzes per reading method. The quizzes for each method were combined for total of 8 possible points per method. As seen in Figure 2, the participants scored better on reading comprehension when reading through paper-based texts rather than on an e-reader system.

A $\chi^2$ McNemar Test for significance of change was calculated for each of the two hypotheses. First, the number of times participants used reading resources with print materials and with electronic materials was analyzed. Results indicated that there was a significant difference between the number of reading resources reported for print and electronic sources, $\chi^2 (1, n = 3) = 7, p < .05$. More resources were used when participants were reading with the
electronic sources than when they were reading print materials even though the same kind of resources were made available for both conditions.

![Reading Resources Used](image)

*Figure 1. Comparison of resources used between the two reading methods*

![Quiz Scores](image)

*Figure 2. Comparison of vocabulary and comprehension quiz scores between the two methods of reading*

Secondly, the frequency of the number of correct quiz responses for those using print books was compared to the number of correct quiz responses for those using electronic books. Results indicated that there was not a significant difference between the number of correct responses reported for print and electronic sources, $\chi^2 (1, n = 3) = .47, p > .05$. While there does seem to be variations in quiz scores, these differences are not significant. Similar to the findings of Grimshaw et al. (2007), these results do not indicate a change in comprehension based on reading source.

**Conclusions and discussion**

The results support the hypothesis that children accessed reading support resources (e.g., a dictionary) more frequently while using an electronic reader. However, the results do not reflect the hypothesis that an e-reading method increases children’s reading comprehension.
The researcher noted that the participants’ reading time was consistently longer when reading on the iPad compared to the print-source. This observation could either challenge the understanding that children born in the new generation are able to automatically adapt to electronic use, or arise multiple other explanations. It may also support that attempting to capture the detail available while screen reading slows down overall reading fluency. One cause could be supported by Grimshaw et al.’s (2007) study, saying that kids are distracted by technology. Another explanation may be that since this was the participants’ first time utilizing the e-reader, the initial “getting comfortable” with a new device and playing with iPad’s features (i.e., automatic page turning by the flicking one’s finger over the page). On the other hand, it could be that the technology enhanced the children’s engagement when reading due to the higher-grade images and electronic front, thus causing them to take longer to complete the required readings. The positive feelings associated with interacting with digital text may have increased the overall motivation to engage in the required readings.

This study differs from Grimshaw et al.’s (2007) study in that the participants of the present study were not able to refer back to the story when taking comprehension quizzes. Within Grimshaw et al.’s (2007) study it cannot be stated with any certainty whether the children were simply turning the pages during the reading time, or actually comprehending the story while reading. Allowing the child to refer back to the story while testing reading comprehension could contribute to the lack of significant differences found between reading books in paper versus electronic formats. For that reason, the present study was concerned with measuring reading comprehension absorbed by the reader while reading the story without the opportunity to refer back to the text.

Notably, each participant’s reading experience was more enjoyable when reading via electronically verse paper print, as reported by the participants themselves. This data was gathered by asking the participants to provide feedback about their reading experiences using the two reading methods. They were asked to state which reading method they liked better, and then to write about why one method was preferred over the other. Each participant selected the iPad over the print book. The following reasons were a few among the responses provided by the participants: The iPad is easier to hold, liked the option to change the screen’s contrast to lighter or darker, enjoyed turning the pages, liked using the electronic book-marker, could understand the story better, and liked the iPad better because the paper book hurts when trying to hold the pages.

Pedagogically, the logical conclusion is that the evidence from this research supports the use of e-texts in reading groups and in the classroom. This is particularly true for students who might not enjoy reading a typical paper-based book or in cases where it is advantageous for a student to access resources. While there is no improvement in comprehension scores, it is important to note that there is no reduction in scores. Further, participants do have an increased use of resources. This could be attributed to any number of items (such as ease of use, enjoyment of use, lack of peer involvement or input on the use of resources). Finally, the participants reported that reading on the iPad was more “fun.”

A limitation of this study was the reduced number of participants. Further research is needed to investigate the effects of this study for students outside the particular eligibility criteria for this study. That is, students from different socioeconomic statuses, age range of students, and students with identified reading or communication delays or disorders.

Based upon the present study’s findings, a question was raised as to why the participants reading comprehension scores did not remain consistent between the two reading sources and why more reading resources were accessed with digital texts. A further study comparing the reading comprehension of longer books which would require longer periods of interaction with the digital texts than this study would provide greater support for the influences of technology on the processes of reading comprehension. Special attention should be given to the manipulability and ease of technological resources available to support reading comprehension (e.g., thesaurus, dictionary, pronunciation) compared to traditional paper-based supports.

References


Appendix A

Parent Questionnaire

Today’s Date:

1. age between 7.0 and 8.11 years: Y N birth date:
2. gender: M F
3. enrolled in second grade: Y N
4. history of special services (speech therapy, use of resource classroom, learning disability): N Y
5. recommended for special education: N Y
6. visual impairment (or glasses): N Y
7. socioeconomic status: Low Middle High
8. English primary Language spoken at home: Y N
9. access to computer at home: Y N
10. access to video games at home: Y N
11. read the story “Nate the Great and the Stolen Base” by Marjorie Sharma: Y N
12. read stories from the “Nate the Great” book series: Y N
13. read the story “Cam Jansen and the Mystery of the U. F. O.” by David Adler: Y N
14. read stories from the “Cam Jansen” book series: Y N
## Appendix B

*Format of reading sessions: order and time span of each activity*

<table>
<thead>
<tr>
<th>Activity</th>
<th>Time span measured by minute</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Print book: reading session I</strong></td>
<td></td>
</tr>
<tr>
<td>get-to-know each-other game</td>
<td>15</td>
</tr>
<tr>
<td>independent reading; pages 7-17</td>
<td>15</td>
</tr>
<tr>
<td>snack break</td>
<td>15</td>
</tr>
<tr>
<td>independent reading; pages 18-29</td>
<td>15</td>
</tr>
<tr>
<td>comprehension quiz</td>
<td>10</td>
</tr>
<tr>
<td>fun activity: make “octopus pops”</td>
<td>20</td>
</tr>
<tr>
<td><strong>Print book: reading session II</strong></td>
<td></td>
</tr>
<tr>
<td>independent reading; pages 30-39</td>
<td>15</td>
</tr>
<tr>
<td>snack break</td>
<td>20</td>
</tr>
<tr>
<td>independent reading; pages 40-48</td>
<td>15</td>
</tr>
<tr>
<td>comprehension quiz</td>
<td>10</td>
</tr>
<tr>
<td>fun activity: scavenger hunt</td>
<td>30</td>
</tr>
<tr>
<td><strong>E-book: reading session III</strong></td>
<td></td>
</tr>
<tr>
<td>independent reading; pages 13-28</td>
<td>15</td>
</tr>
<tr>
<td>snack break</td>
<td>20</td>
</tr>
<tr>
<td>independent reading; pages 29-41</td>
<td>15</td>
</tr>
<tr>
<td>comprehension quiz</td>
<td>10</td>
</tr>
<tr>
<td>fun activity: make alien goo</td>
<td>30</td>
</tr>
<tr>
<td><strong>E-book: reading session IV</strong></td>
<td></td>
</tr>
<tr>
<td>independent reading; pages 42-55</td>
<td>15</td>
</tr>
<tr>
<td>snack break</td>
<td>20</td>
</tr>
<tr>
<td>independent reading; pages 56-65</td>
<td>15</td>
</tr>
<tr>
<td>comprehension quiz</td>
<td>10</td>
</tr>
<tr>
<td>reading method questionnaire</td>
<td>10</td>
</tr>
<tr>
<td>fun activity: create paper U. F. O.</td>
<td>30</td>
</tr>
</tbody>
</table>
Appendix C

Quiz #1

1. What does mascot mean? “They are the team’s mascots.”
   a. a wild animal
   b. a good luck charm
   c. a Halloween costume
   d. a type of food

2. What does sorted mean? “I, Nate the Great, moved things, plied things, and sorted things”
   a. to put things in the right place
   b. to make things messy
   c. to throw away things
   d. to tear things

3. Where is the first place Nate looked for the missing/stolen octopus?
   a. In a tree
   b. at the playground
   c. In the Oliver’s trashcan
   d. on Oliver’s bookshelf

4. Rosamond thinks her cats could find the octopus if it was not made of plastic. Why?
   a. Her cats do not like plastic things
   b. plastic octopuses are harder to find
   c. a real one would smell like fish and cats eat fish
   d. it would be easier to find swimming in the ocean

Quiz #2

1. What does snatch mean? “Did Oliver see the snatch?”
   a. to hit a homerun
   b. a kids t.v. show
   c. to steal something
   d. a type of insect

2. What does peered mean? “Then I peered behind the bookcase.”
   a. to look closely
   b. to walk slowly
   c. to look quickly
   d. to sit for a long time

3. Who did Nate think stole the base first?
   a. Rosamond’s cats
   b. an older kid at school
   c. the other team
   d. Annie’s dog, Fang

4. At the end of the story when Nate was up to bat, why did he hope to strike out?
   a. He was scared of Fang, who was on the third base
   b. He wanted to play with Fang and his bone instead
   c. to look for another mystery to solve
d. He did not want to lose the baseball after he hit it

**Quiz #3**

1. What does *posed* mean? “I can’t take a posed picture,” Eric told her.
   a. to be famous
   b. to pretend or be fake
   c. to be old
   d. to be creative

2. What does *squaw* mean? “We’ve seen a red-backed sandpiper, a bufflehead, an old-squaw, and now this.”
   a. a Native American woman
   b. a type of snake
   c. a dinosaur fossil
   d. a type of vegetable

3. Why did the reporter walk away before Cam finished talking to her?
   a. U. F. O. stories scare the reporter
   b. The reporter did not like the cat liking her
   c. Cam was not telling the truth
   d. The reporter saw a UFO in the sky

4. Why did they name the cat Neptune; an outer-space name?
   a. the cat likes stars at night
   b. the cat looks like a planet
   c. they thought the cat came from the U.F.O.
   d. they found the cat at the same time they saw the U. F. O.

**Quiz #4**

1. What does *stroked* mean in the sentence, “She stroked the back of her neck”?
   a. to hit something hard
   b. to gently move your hand across something
   c. to paint
   d. to give a kiss

2. What does *shield* mean in the sentence, “Cam held her hand over her eyes to shield them”?
   a. a type of hat
   b. a pair of glasses
   c. a piece of armor carried on arm
   d. a way to protect something

3. Why did Neptune move towards the end of the tree branch when she saw Cam?
   a. Cam will give her food if Neptune shakes the branch
   b. Neptune was scared
   c. They were playing the game, hide-and-go-seek
   d. Neptune lives in the tree

4. What did the kids use to make the UFOs?
   a. Metal
   b. Candy, tape, and paper plates
   c. Balloons and flashlights
   d. They were real UFOs from space
Design of Online Report Writing Based on Constructive and Cooperative Learning for a Course on Traditional General Physics Experiments

Hao-Chang Lo
Department of Digital Content and Technology, National Taichung University of Education, 140 Min-Shen Road, Taichung 40306, Taiwan // simonlo@mail.ntcu.edu.tw

(Submitted September 09, 2011; Revised February 18, 2012; Accepted February 29, 2012)

ABSTRACT
The objective of this study was to develop an online report writing activity that was a constructive and cooperative learning process for a course on traditional general physics experiments. Wiki, a CMC authoring tool, was used to construct the writing platform. Fifty-eight undergraduate students (33 men and 25 women), working in randomly assigned groups of 2 to 3 members, participated in this course. Both quantitative and qualitative data, including assessments of reports, questionnaires, interviews, and records of discussion on wiki, were collected and analyzed to investigate the course design of online report writing. Results showed that students recalled, discussed, searched for, and integrated auxiliary information, reflected on the experiment, and shared meaning in the process of online writing. Evidence of positive interdependence, promotive interaction, individual accountability, social skills, and group processing proved that students worked cooperatively to accomplish shared learning goals. A higher average score of online writing than that of traditional paper writing indicated that students wrote online reports better, compared to the traditional method. Student participants and the instructor in the course responded positively when they were questioned on their perceptions of the Wiki-based report writing.

Keywords
Constructive learning, Cooperative learning, General physics experiment, Laboratory report, Online writing

Introduction
The educational laboratory has been a common feature of introductory courses since the 1800s, and has received emphasis during reforms in the 1960s (Blosser, 1983). Traditional laboratories/cookbook experiments are often derided in the research literature as “old-style labs” because the reform movement in science education has led laboratory work to be an inquiry and problem-solving process. However, it seems that even today cookbook experiments are used in many physics laboratories. The reason cookbook laboratories are popular may be because they directly assist undergraduate students in learning how to conduct different experiments, to work in “real” science laboratories in large-scale classrooms with few instructors.

Writing a laboratory report, a type of science writing, plays an essential role in laboratory activities. Experiments may aid not only in the development of conceptual thinking and imagination, but also in fostering scientific practices in the classroom. It also provides a basic means in science for evaluating raised alternatives as possible solutions (Kreitler & Kreitler, 1974). The principal goal of the laboratory report is to document learners’ findings and communicate their significance. A good laboratory report does not only present data, but also demonstrates student comprehension of the concepts behind the data.

Laboratory report writing serves a specific purpose in science learning, and cookbook laboratories are used in numerous courses on general physics experiments. Developing a laboratory report writing activity to foster students’ understanding of experiments is essential in traditional general physics laboratory work.

Constructivism asserts that learning is an active, interpretive, and iterative process (Tobin, 1990). Learning is contextualized, and learners construct knowledge by solving genuine and meaningful problems (Brown, Collins, & Duguid, 1989; Polman, 1999). Students need opportunities to negotiate their understanding, both within a social setting and the current knowledge of the scientific community (Prawat, 1989). Students share multiple perspectives and generalize their understanding and knowledge to enable application in different contexts when they articulate their knowledge to one another (Collins, 1991). Articulation involves students thinking of their actions and providing reasons for their decisions and strategies, making their tacit knowledge more explicit or overt (Wilson & Cole, 1996). Articulation can be achieved in various other means, including working in groups, discussing and debating...
issues, reporting, presenting findings, and negotiating and defending knowledge acquired through learning environments. Reflection, which is a closely related idea, is proposed as well (Wilson & Cole, 1996). Reflection includes the process of analyzing and making judgments on occurrences to provide new meaning to a situation.

For helping students construct scientific knowledge when they write laboratory reports after class, writing activities should provide students with the opportunity to recall the experimental process, to communicate and negotiate the understanding of scientific activity, and to think reflectively when engaging in their writing. Johnson and Johnson (1991) perused scientific journals and showed that most scientific research is conducted by groups of scientists. This indicates that laboratory report writing, as a part of scientific research, should be finished cooperatively by student groups, and writing should be a cooperative learning process.

Johnson and Johnson (1984) claimed that positive interdependence, promotive interaction, individual accountability, group processing, and social skills are the five essential elements of cooperative learning. Positive interdependence means that students must participate fully and exert effort within their group, and each group member has a task/role/responsibility; therefore, they must believe they are responsible for their individual learning and the learning of their group. Promotive interaction means that students explain to each other what they are learning, and that they assist one another with understanding and completing assignments. Individual accountability means that each student is accountable for their own learning and work, thereby eliminating “social loafing.” Group processing means that groups must occasionally assess their effectiveness and decide possible approaches for improvement. Social skills, such as effective communication and interpersonal and group skills must be taught for successful cooperative learning to occur (Brown & Ciuffetelli Parker, 2009).

Generally, undergraduate students do not write their laboratory reports in class but finish report writing at home, in accordance with the formal modeled report format. Approaches for cooperative working address an existing problem when students write lab reports. Computer-mediated communication (CMC) provides a solution for the limitation of traditional laboratory report writing.

With widespread Internet usage, CMC use includes synchronous and asynchronous modes and is a part of students’ daily lives. Numerous open source CMCs are available, and writing using the CMC tool is similar to writing in a Word processor because it does not require prior knowledge of HTML codes or programming languages (Richardson, 2006). Wiki, one of the open-source CMCs, provides a Web space for social interaction and collaboration (Godwin-Jones, 2003; Su, 2005). It enables communities to write documents collaboratively, using a simple markup language and a Web browser. The permanent retention of each interaction of posts in a wiki provides learners with the opportunity to explore the evolution of any wiki page and, if deemed appropriate, replace the current version with a previous iteration. A wiki allows all users to edit any page or to create new pages within the wiki site. Thus, a contribution is not a comment or response (as it might be in a blog), but an alteration to the previous contribution. This means that a wiki-based text is in a constant state of potential collaborative change (Kessler, 2009).

Researchers have developed several successful methods of collaborative learning that can be used to integrate science, reading, and writing (Nesbit & Rogers, 1997). Because constructivist ideas have gained support, science writing has been perceived increasingly as an approach to bridge prior knowledge with new learning, to build explanations, and to enhance comprehension of information from sources such as class discussions, laboratories, or textbooks (Santa & Havens, 1991; Prain & Hand, 1996). Keys, Hand, Prain, and Collins (1999) presented a new heuristic tool called science writing heuristic (SWH) to promote learning from laboratory activities in a secondary school science course. The purpose of SWH was to enhance student thinking, discussion, and metacognition when students engaged in science learning (Keys, 2000). They claimed that students developed a personalized understanding of scientists’ work, the nature of evidence, and collaboration in science by using SHW in their writing tasks in the classroom.

The objective of this study was to develop an online report writing activity that was constructive, and a cooperative learning process for a course on traditional general physics experiments. To evaluate the course design of online report writing, four questions are examined: (1) Was the online writing activity constructive for science learning? (2) Did students work cooperatively to accomplish shared learning goals? (3) Was the achievement different between online writing and traditional writing? (4) What did participants think of online writing?
Methods

Participants

Participants comprised 58 freshmen (33 men and 25 women who enrolled in a course on general physics experiments and the instructor who taught the course. Students had taken general physics credits, and the instructor, having graduated from the Institute of physics, had taught the course on general physics experiments for over ten years.

Instructional design

There were ten experiments (experiments 1-10, one experiment for three hours once per week), including (1) fundamental measure, (2) Newton's second law of motion, (3) linear acceleration caused by gravity, (4) interference of waves, (5) diffraction, (6) Human Eye Model, (7) Volt-Ohm-Milliammeter, (8) oscilloscope and Lissajous pattern, (9) transformer, and (10) the mass-to-charge ratio.

Students were randomly assigned to 20 groups (2-3 for each group). For research purposes, students had to experience two different types of writing, to perceive the difference between online writing and traditional writing. Students experienced online writing first (Week 1 - Week 5), and subsequently, traditional writing (Week 6 - Week 10) in the same group assignment; that is, students finished reports using the Internet in the first 5 weeks, including pre-lab reports and lab reports. To prepare students for online report writing, the instructor demonstrated the writing platform and the manner in which to write a cooperative report using the platform before starting the course. For Weeks 6-10, students conducted their work in the traditional manner of most general physics laboratories. They were asked to submit pre-lab reports on paper before the experiment and lab reports on paper after lab work, without guidance, discussion, writing using the online writing platform, and consequently, the instructor did not participate in the process of report writing.

Methods of implementing online report writing

For the purpose of implementing constructive and cooperative online writing, Wiki was used to build the writing platform, and the concept of Keys' work of SWH was adopted when the course was developed. A template called online report writing guideline (ORWG) was designed to guide students to complete their experiment reports stepwise. Instead of the traditional standard report format, ORWG presents eight questions (Table 1), prompting students to think and write.

Table 1. Comparison of ORWG with standard report format

<table>
<thead>
<tr>
<th>Student Template</th>
<th>Standard Report Format</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-lab report</strong></td>
<td></td>
</tr>
<tr>
<td>Q1. What is the objective of this experiment?</td>
<td>Title, purpose</td>
</tr>
<tr>
<td>Q2. What theory/rules are involved?</td>
<td>Theory, rules, hypothesis</td>
</tr>
<tr>
<td>Q3. What/how should we conduct the experiment?</td>
<td>Apparatus, materials, and planned procedure</td>
</tr>
<tr>
<td><strong>Lab report</strong></td>
<td></td>
</tr>
<tr>
<td>Q4. What have we conducted? What’s difference between what we did and what was planned in the pre-lab report?</td>
<td>Procedure</td>
</tr>
<tr>
<td>Q5. What was observed in the experiment?</td>
<td>Data, observation</td>
</tr>
<tr>
<td>Q6. How do we analyze the experimental data and what are the results?</td>
<td>Results (texts, graphs, tables)</td>
</tr>
<tr>
<td>Q7. What are the findings and how do we explain them?</td>
<td>Finding</td>
</tr>
<tr>
<td>Q8. Do the findings agree with the theory (or rules)? If not, can we explain why?</td>
<td>Test, Conclusion</td>
</tr>
</tbody>
</table>

Prior to the laboratory, students write the title/purpose of their experiment (Q1: What is the objective of this experiment?), the theory (or rules) to be tested (Q2: What theory/rules are involved?), and the apparatus/materials, subjects, equipment, and procedure according to the textbook and, if necessary, information they search (Q3: What/how should we conduct the experiment?). After the experiment, students recall their actions and compare them...
to what they had written in the pre-lab report (Q4: What have we done? What’s the difference between what we conducted and what was planned in the pre-lab report?), observation and data (Q5: What was observed in the experiment?), results of data analysis and presentation (Q6: How do we analyze the experimental data and what are the results?), findings (Q7: What are the findings and how do we explain them?), and draw conclusions (Q8: Do the findings agree with the theory (or rules)? If not, can we explain why?). Students are guided to recall, reflect on the experiment, and they also have an opportunity to negotiate, discuss, and share interpretations during the online writing process.

Students logging in using their student number are brought to the homepage display of the online report writing, as shown in Figure 1. Operating instructions, links, navigation, search, toolbox, and ORWG are included in the homepage.

When students click on the community icon on the navigation, they can choose the experiment they want to enter and what group they belong to (Figure 2). For example, Figure 3 shows a lab report writing page when students (Group 1) write Experiment 4. They write the report collaboratively, including the pre-report and post-report, by using the ORWG. Students can engage in discussions on the Discussion page (Figure 4) after they click the Camel Case link.
Data collection and analysis

Both qualitative and quantitative data, including report assessments (A), a questionnaire (Q), interview (I), and records of discussion on wiki (D), were collected and analyzed to investigate the design of online report writing. Figure 5 shows the relation between the data resource and research questions. For example, to examine participant perceptions on the differences between online writing and traditional writing on paper, we conducted an interview with the instructor and provided a questionnaire to students at the end of the course. Furthermore, student discussions on the wiki were also analyzed.

Instruments

Two instruments, a scoring rubric and a questionnaire, were developed for data collection. A scoring rubric (Appendix 1), including criteria and its description, was developed using a four-round Delphi process to score student reports. A pilot study showing evaluators had a highly significant correlation ($\omega = .930; p < .001$) when using
the scoring rubric (Lo, 2010). Seventeen criteria are in the scoring rubric, including (1) title, (2) purpose, (3) hypothesis, (4) principles, (5) procedures, (6) methods, (7) observation, (8) graphs and tables, (9) data presentation, (10) data integration and analysis, (11) conclusion, (12) report structure, (13) overall presentation, (14) understanding of experiment, (15) auxiliary data, (16) theory discussions and application, and (17) error analysis. A questionnaire was developed to investigate student experiences of online report writing. Students answered questions regarding their thoughts on online report writing with wiki usage.

![Figure 5. Illustration of the relation between data and research question](image)

**Results and findings**

*Construction of scientific understanding in the process of online writing*

Students presented a question related to the experiment when they wrote the pre-experiment report, as stated in the following examples:

“What is a vibration generator?” (D S36 2009/11/23 10:25)

“It seems that the periodic wave on a string will be produced by it. Anyway, we will know this in the class the day after tomorrow.” (D S38 2009/11/24 08:47)

Students shared the searched information with group members by using the hyperlink, as stated in the following examples:

“Here is Ohm's law published by Georg Ohm in the book Die galvanische Kette, mathematisch bearbeitet (1827). It states that…” “You can see more on the following link: http://en.wikipedia.org/wiki/Ohm's_law.” (D S24 2009/11/23 10:12)

Students recalled and reflected on the experimental process, as stated in the following examples:

“Why did the experiment fail when we conducted it the first time?” (D S16 2010/3/14 22:27)

“It seems that the mass cannot touch the ground before the trolley reaches the pulley.” (D S17 2010/3/14 22:33)

“Yes! We must choose a length of string, such that the mass does not touch the ground until the trolley nearly reaches the pulley when we want to explore the relationship between the magnitudes of the external force and the resulting acceleration.” (D S31 2010/3/14 23:39)

“Why?” (D S16 2010/3/14 24:08)

“The magnitudes of the external force are zero when the mass touches the ground.” (D S31 2010/3/14 25:25)

“I get it.” (D S16 2010/3/14 28:14)
Students discussed the experiment they had conducted. Moreover, they found that they were making errors in the experiment, as stated in the following examples:

“It seems that we did it wrong when we connected the parallel circuit. Why did we not detect it then?” (D S51 2009/12/09 20:31)
“What is the next step? Should we give up on this report?” (D S52 2009/12/09 21:10)
“We can invite the instructor to discuss it and ask him to give us another chance to do it again.” (D S51 2009/12/09 21:18)

They analyzed the observed data and compared the empirical results with a predicted one, as stated in the following examples:

“Does anyone remember what the value of the hanging mass was?” (D S19 2010/3/10 11:43)
“8.27 g.” (D S20 2010/3/13 8:09)
“Is the predicate value of acceleration 40.877 or 0.4877? Can anyone figure this out?” (D S21 2010/3/13 10:18)
“8.27*9.8/198.27=0.40877. Is this correct?” (D S22 2010/3/13 10:27)
“I don’t think so. The gravitational acceleration, g, is 9.8 m/s², so we should use Kg as the unit of measurement.” (D S20 2010/3/13 10:30)
“0.00827*9.8/0.19827=0.40877. Is this correct?” (D S22 2010/3/13 22:20)
“The observed data a=0.38 m/s². How do we explain this? Can we claim that Newton's second law of motion is wrong?” (D S19 2010/3/14 12:15)
“The force of friction is not considered. We could try to discuss the force of friction.” (D S22 2010/3/14 12:20)
“I think Newton's second law of motion will work if the force of friction and experimental error are considered.” (D S20 2010/3/16 23:58)

Students linked the experiment with what they had learned in general physics, as stated in the following examples:

“The objective is to measure the velocity of waves by the tension in the string and the mass per unit length of the string.” (D S38 2009/11/25 10:14)
“I remember that this equation had been taught and we tested it in the general physics course.”
“Yes, we did.” (D S40 2009/11/25 11:08)

They also occasionally discussed the theory or rules related to the experiment, as stated in the following examples:

“I found an error in the equation...” (D S39 2009/11/25 11:19)
“I think that E squared in the equation is wrong. Is it not?” (D S18 2009/11/25 21:16)
“Yes. It is a mistake. I checked that E is not squared.” (D S19 2009/11/26 11:09)

The ORWG did not only help students finish their reports stepwise, but also enhanced students’ reflections on what had they conducted in the cookbook laboratory experiments. A student stated on the wiki page:

“At first, I did not know why the guideline asked me to do what I had I done in pre-lab reports writing. I finally realized that I would know the differences between what we had done in experiments and what we had written/predicted in pre-lab reports.” (D S37 2010/4/25 11:28)

The evidence to prove that online writing process is a cooperative learning

Students encouraged group members to engage in teamwork or reminded group members to prepare for the experiment, as in the following examples:

“Please participate in the discussion when you are free, so that we are good partners in Group 4.” (D S1 2010/3/9 21:40)
“To all members in Group 7, it is good for you to enter the discussion about report writing, and the instructor will score it.” (D S2 2010/3/8 22:31)
“They said that it will be hard work tomorrow. Let’s study the experiment hard in advance.” (D S32 2010/3/23 21:01)

Students provided group members with support and pursued the instructor for support to finish the report when they engaged in their work, such as in the following examples:

“Thank you for your help. I failed to finish the figure several times. I do not know whether it was an issue with my computer.” (D S46 2009/12/16 13:00)

“You are welcome. Don’t you remember that we are members of the same team?” (D S44 2009/12/17 12:27)

“Instructor X, I could not draw a table on the wiki, so I finished it in Word and then posted it on the wiki.” (D S28 2009/12/02 14:02)

A new style of learning and unfamiliar writing platform overwhelmed students when they began writing on the wiki. They communicated, negotiated, and practiced to finish their tasks effectively, such as in the following examples:

“I have something to say. Be sure to space the word, and then a broken blue line will appear.”

“Dear my group member, thank you for your suggestion. I will do it.” (D S54 2009/11/25 00:05)

“It perplexed me because of my unfamiliarity with the wiki operation.” (D S53 2009/11/25 16:33)

“Practice makes perfect. I finished it faster than last time.” (D S50 2009/11/19 10:21)

“Maybe we should meet online at a certain time every week, or our work might seem to lack effectiveness.” (D S31 2009/11/08 11:15)

“I agree. It is good for our assignment.” (D S16 2009/11/08 11:34)

“I listed a task assignment of who will be an editor for each experiment. I think it will help us finish effectively.” (D S51 2009/10/17 10:29)

“It’s a good idea that there will be a leader for each report.” (D S52 2009/10/17 11:05)

**Differences between online writing and traditional writing achievement**

To compare the performances of online writing with traditional writing, all reports were scored according to the scoring rubric developed in this study. The statistical result is shown in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Online writing</th>
<th>Traditional writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>69.82</td>
<td>65.68</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12.42</td>
<td>13.22</td>
</tr>
<tr>
<td>Value of t</td>
<td>2.283*</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at p < .05

**Participant Perceptions on the differences between online writing and traditional writing**

Participants would discuss/write/assess report online without limitations of time and space.

“We can write reports and discuss anything about the experiments online if we want.” (Q S27 2010/01/15 11:13)

“I participated in the students’ report writing and examined their reports through the Internet when I was available.” (I Instructor 2010/01/28 11:10)

Writing reports online was a new and interesting approach and a challenge to students. However, it took time to familiarize with the writing platform, but the ability to use information technology would be enhanced after learning.

“Cool! Writing reports online.” (D S18 2009/10/05 21:53)

“It is refreshing to know that we will finish our report online.” (D S20 2009/11/05 21:55)

“At the beginning of online writing, it was a challenge for me because of my poor ability to use my computer. However, it seems easy for me to write and to communicate online now.” (Q S46 2010/01/15 11:51)
The ORWG template developed in this study facilitated student observations on the experimental process and helped write a more detailed report. According to the wiki guidelines, students would reflect on the experiments and finish reports stepwise. Because the wiki page records all revision history, students were motivated to participate in discussions and report writing.

“At first, I did not know why the guideline asked me to do what I had I done in pre-lab reports writing. I finally realized that I would know the differences between what we had done in experiments and what we had written/predicted in pre-lab reports.” (D S37 2010/4/25 11:28)

The instructor knew more of an individual’s contribution of report writing through the wiki page and could assess student reports more adequately.

“I, as an instructor, feel that the best advantage of online writing is that I can perceive each student’s contribution to the report based on records on the writing platform. Students are scored according to their individual effort.” (I Instructor 2010/01/28 11:18)

Participants thought that online report writing is “convenient” and “saves energy and reduces carbon emissions.”

“It is a good idea to write reports online because of the decrease in wasted powdered carbon and paper.” (D S18 2009/10/05 21:53)

“I will not have to take hefty paper reports with me if online report writing is conducted.” (I Instructor 2010/01/28 11:29)

Discussion

Students have the opportunity to negotiate their understandings both within their social setting and against the current knowledge of the scientific community (Prawat, 1989). This type of report writing can be perceived as an approach to bridge prior knowledge with new learning, build explanations, and enhance understanding of information from sources such as class discussions, laboratories, and textbooks (Santa & Havens, 1991; Prain & Hand, 1996). Students also articulated and reflected (Oliver, Herrington, & Omari, 1996; Wilson & Cole, 1996) in the process of science writing. Hoadley and Linn (2000) claimed that asynchronous peer discussions can be designed to enhance the cohesive understanding of science.

Bosworth (1994) indicated that it is an important strategy to develop a healthy atmosphere within small groups to encourage students to participate willingly and question each other. Furthermore, a good interaction helps students develop a sentiment that each team member is responsible for contributing to the group. Group writing had a positive interdependence because the final products depended on contributions from all group members, and students believed that there was value in working with other group members, and that both individual learning and work products would be enhanced as a result of the collaboration. According to ongoing conversations, dialogues, exchanges, and support, promotive interaction occurred in the task. This also indicated that the instructor could assess each student based on records on the wiki page, and returned the results to the groups and individual students to ascertain contributions by each member (individual accountability). Communicating and negotiating on the wiki page were essential social skills during group tasks, and the dialogue showed that students achieved their objectives by placing effort at the beginning of their work. Students also negotiated with group members to improve their work in the writing process (group processing). The student statement examples indicate that online report writing is a cooperative learning process (Brown & Ciuffetelli Parker, 2009; Johnson & Johnson, 1984).

Feedback from participants indicated that they thought the differences between online writing and traditional writing were that online writing provided participants with the opportunity to discuss, write, and assess reports without limitations of time and space, that the ORWG helped student observations on the experimental process and write more detailed reports, and that writing online could save energy and reduce carbon emissions, but it took time to familiarize themselves with the writing platform. Because of the record of revisions in the wiki, students were motivated to participate in discussions and report writing, and the instructor could determine students’ individual
contributions to the report. Students thought of online writing as a new, interesting, and challenging work that could be completed stepwise according to the guidelines.

The statistical result of writing achievement, as shown in Table 3, shows a higher average score of online writing than that of traditional paper writing \((t = 2.283, p < .05)\). This indicates that students wrote better reports online than they did using the traditional approach in the same group assignment. Kagan (1992) indicated that one of the approaches that shown a positive result in boosting student writing skills is the incorporation of cooperative learning. In this study, the online writing platform provided an environment for cooperative learning, the discussions and writing of which continued without limitations of time and space. That might be the reason why students wrote better reports online than they did using the traditional approach in the same group assignment. Furthermore, the ORWG, which was provided on the wiki page, facilitated student observations on the experimental process, and students would finish more detailed reports stepwise, which would also enhance students’ writing performance.

**Future work**

Despite the encouraging results of this study regarding the positive effect of incorporating online report writing into a course on general physics experiments, future research is required in a number of directions. First, the direct evidence on how online writing facilitates the performance of students’ report writing must be clarified. Second, the study results can encourage instructors to attempt to incorporate online report writing into other traditional basic experiment courses such as chemistry and biology. Further studies should be undertaken to evaluate the effects of using other free CMC tools to support experiment courses. Third, improved designs for writing platforms are essential to popularize online report writing, such as embedded tools for drawing figures and tables, and a friendly writing environment that students can work in with less training. Finally, an online assessment system for online writing is necessary for when online report writing is widely implemented.

**Conclusions**

This study presented a design of online report writing, which engaged students to finish their general physics experiment reports constructively and cooperatively. Results showed that students recalled, discussed, searched, and integrate auxiliary information, reflected on experiments, and negotiated and shared interpretations throughout the process of online writing. This means that the online report writing activity provided students with the opportunity to construct an understanding and knowledge of the scientific experiment. Evidences of positive interdependence, promotive interaction, individual accountability, social skills, and group processing proved that students worked cooperatively to accomplish shared learning goals. Students wrote better reports online than they did using the traditional approach when they worked with the same group members. Students and the instructor who participated in the course responded positively when they were questioned on their perceptions of the Wiki-based lab report writing.

This suggests that the benefits of using CMC tools in report writing are to create an environment where students can write reports cooperatively, and without limitations of time and space, to improve the writing process as an approach to constructive learning, and especially to provide instructors with the opportunity to engage in students’ report writing process. The design of online report writing-based constructive and cooperative learning for courses on traditional general physics experiments is good for both teaching and learning. Furthermore, instead of using paper, online reporting is an implementation entailing the concept of saving energy and reducing carbon. Although this study has its limitations, we hoped that it can serve as a foundation for future research in incorporating CMC tools into science writing.

**Acknowledgements**

This work is supported by the National Science Council, Taiwan, Republic of China under contract NSC98-2511-S-040-001. We would like to extend our heartfelt gratitude to instructor Lee for assisting in the implementation of the teaching for the research and Dr. Wang for English writing consultation.
References


## Appendix 1

### Laboratory Report Scoring Rubric

<table>
<thead>
<tr>
<th>Category</th>
<th>Grading Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td>A (4-5) The title is complete and precise.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The title is not quite understandable.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) No title is shown.</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>A (6-8) The purpose is fully described.</td>
</tr>
<tr>
<td></td>
<td>B (3-5) The purpose is not quite understandable.</td>
</tr>
<tr>
<td></td>
<td>C (0-2) There is no purpose, or the purpose is wrong.</td>
</tr>
<tr>
<td><strong>Hypothesis</strong></td>
<td>A (6-8) The description and the result do not match.</td>
</tr>
<tr>
<td></td>
<td>B (3-5) The hypothesis is not completely logical.</td>
</tr>
<tr>
<td></td>
<td>C (0-2) There is no hypothesis, or the hypothesis is wrong.</td>
</tr>
<tr>
<td><strong>Principles</strong></td>
<td>A (4-5) The principles are described clearly.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The principles are slightly mentioned, but not fully described.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) There is no related principle or it is inaccurate.</td>
</tr>
<tr>
<td><strong>Procedures</strong></td>
<td>A (4-5) Every procedure is clearly and fully described.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The procedures are described roughly, but lack clarity.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) A description for the procedure is absent.</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>A (4-5) The methods are clear and purpose-related.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The methods are roughly described, but not purpose-related.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) A description or principle is absent.</td>
</tr>
<tr>
<td><strong>Observation</strong></td>
<td>A (6-8) It is well-organized with accurate observations and clear explanations or notes.</td>
</tr>
<tr>
<td></td>
<td>B (3-5) It is well-organized, but not fully described.</td>
</tr>
<tr>
<td></td>
<td>C (0-2) No observations have been made.</td>
</tr>
<tr>
<td><strong>Graphs and tables</strong></td>
<td>A (6-8) The graphs and tables are labeled properly, are understandable, and with a complete description.</td>
</tr>
<tr>
<td></td>
<td>B (3-5) The graphs and tables are all complete and clear, but some information partly lacks explanations.</td>
</tr>
<tr>
<td></td>
<td>C (0-2) There is no label or description.</td>
</tr>
<tr>
<td><strong>Data presentation</strong></td>
<td>A (4-5) Results are clearly listed, concise, and well-organized using the most appropriate representation of all data.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) Results are roughly shown, without clear and logical organization.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) Results are not shown, or essential data are missing.</td>
</tr>
<tr>
<td><strong>Data integration and analysis</strong></td>
<td>A (6-8) All data are appropriately integrated and analyzed, and fully noted with details.</td>
</tr>
<tr>
<td></td>
<td>B (3-5) Most of the data are appropriately integrated and analyzed.</td>
</tr>
<tr>
<td></td>
<td>C (0-2) The data are not appropriately integrated and analyzed.</td>
</tr>
<tr>
<td><strong>Conclusion</strong></td>
<td>A (4-5) The conclusion is fully supported by the hypothesis.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The conclusion is roughly drawn based on results, but is not supported by the hypothesis.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) The conclusion has not been drawn.</td>
</tr>
<tr>
<td><strong>Report structure</strong></td>
<td>A (4-5) The structure of the report is appropriate and is complete, with clear titles and subtitles.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The structure of the report is clear but not well-organized.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) Original: Report is not well-organized and lacks a clear structure.</td>
</tr>
<tr>
<td><strong>Overall presentation</strong></td>
<td>A (4-5) The overall report is written in scientific style: clear and concise. It is also logically and concretely presented, with no typos.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The overall report is roughly written in scientific style, but lacks logical and concrete presentation. Some typos are also present.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) The overall report is not logically and concretely written.</td>
</tr>
<tr>
<td><strong>Understanding of experiment</strong></td>
<td>A (4-5) Students can write a clear understanding and the goal in the report.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) Students only write a general idea in the report.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) Students do not write anything based on their understanding in the report.</td>
</tr>
<tr>
<td><strong>Auxiliary Data</strong></td>
<td>A (4-5) Students collected and organized data completely. Data are also clearly presented as the basis of the discussion.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) Students collected and organized data, but they are incomplete.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) Original: Students failed to provide related data.</td>
</tr>
<tr>
<td><strong>Theory discussions and application</strong></td>
<td>A (4-5) The theory discussion and the application are complete.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) The theory discussion and the application are roughly mentioned.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) A theory discussion and/or the application are absent.</td>
</tr>
<tr>
<td><strong>Error analysis</strong></td>
<td>A (4-5) Experimental errors are discussed, and experimental methods or techniques are presented to reduce the experimental errors.</td>
</tr>
<tr>
<td></td>
<td>B (2-3) Experimental errors are mentioned but not discussed.</td>
</tr>
<tr>
<td></td>
<td>C (0-1) Experimental errors are not mentioned.</td>
</tr>
</tbody>
</table>
Meeting the “Digital Natives”: Understanding the Acceptance of Technology in Classrooms

Xiaoqing Gu*, Yuankun Zhu and Xiaofeng Guo

Department of Educational Information Technology, East China Normal University, Shanghai, China, 200062 // xqgu@ses.ecnu.edu.cn // zhyk99@yahoo.com.cn // guoxiaofeng2010@hotmail.com

*Corresponding author

(Submitted September 26, 2011; Revised November 22, 2011; Accepted January 07, 2012)

ABSTRACT

The past few decades have witnessed the rapid development of information and communication technology around the world, as well as continuing efforts to introduce technology into K12 schools. To gauge the success of integrating technology into classrooms, how end users, including teachers and students, accept and use technology while overcoming a new kind of digital gap needs to be investigated. To better assess the integration of technology into the classroom experience, the current study aimed to understand the difference between teachers and students’ acceptance of technology. The participants in this study were chosen through stratified random sampling in Shanghai. A research model with related factors compiled from literature on technology acceptance was used to collect data. The results indicated that the differences between teachers and students with regard to technology lie in how they utilize technology and how important they perceived it to be. The results of this study may help us better understand new millennium learners and provide them proper classroom technology products.

Keywords
ICT-integration, Digital natives, New millennium learners, Digital immigrants, Technology acceptance

Introduction

Information and communication technology (ICT) has been introduced into K12 schools since the late 1990s with the expectation that it can transform education and facilitate students’ learning in China (Gu & Ouyang, 2008; Zhang, Fang, & Ma, 2010) and other countries (Hew & Brush, 2007). Like other IT application areas, the success of technology integration into classrooms depends on how end users, including teachers and students, accept and use technology. A better understanding of the benefits, barriers, and other factors affecting the end users’ acceptance of technology would be helpful in providing them with the proper support, services, and tools.

Studies on technology acceptance and its associated factors in the area of IS and end-user computing have proliferated, and a variety of theoretical models have attempted to explain the determinants of individual acceptance and use of information technologies. In these studies, however, little attention has been paid to end users of technology in classrooms, where the impact of technology depends on both the teachers and students who use it. Between these two groups, a new kind of digital gap is emerging.

Current students are more knowledgeable than their teachers when it comes to ICT. Given this phenomenon, Prensky (2001) coined the terms “digital natives” to describe students and “digital immigrants” to describe teachers. Today’s students have also been called the “Net Generation” (Oblinger & Oblinger, 2005) and “new millennium learners” (Pedró, 2007). These students have been raised in a digital environment that has shaped how they think, behave, and act. Therefore, the nature of technology usage in and out of schools and the acceptance of technology between digital natives and digital immigrants are presumably radically different.

The present study seeks to understand how the two groups of end users differ in technology usage in and out of school, as well as their respective attitudes toward it. Literature related to the success, failure, and barriers to ICT integration from the perspective of the end users was initially reviewed. The constructs from technology acceptance models and IT success literature were then compiled. Empirical data were gathered to find the differences among various situations. Given that the technology integrated into classrooms is designed by teachers for the benefit of students, knowing the difference of technology acceptance among teachers and students could help in the development of classroom technology products that cater to digital natives.
Literature review

Conditions that do not support the integration of ICT can be considered barriers. In their efforts to examine the current barriers related to the integration of technology in the curriculum of K12 schools, Hew and Brush (2007) analyzed existing empirical studies of technology integration from 1995 to spring 2006 in the United States and other countries. Of the six categories of barriers that were examined, at least two were related to teachers’ behavior: the lack of specific knowledge and skills about technology integration, and attitudes and beliefs toward technology. The link between these types of barriers was constructed as well. For instance, teachers’ attitudes and beliefs are affected by their knowledge and skills. Bingimlas (2009) similarly reviewed the literature of technology integration in science education. He found that although teachers had a strong desire for integrating ICT into education, they encountered many barriers, one of which was the lack of confidence and competence, or having negative attitudes and inherent resistance. Again, the barriers are somehow inter-related. These studies indicate that teachers’ experiences, attitudes, and competence are vital to the success of ICT integration.

Like the role teachers play in the integration of technology, students are also actors in ICT integration in the classroom. Academic commentators argue that the computer is the “children’s machine” (Papert, 1993), and that the ICT integration in classrooms is done for the kids (Selwyn, 2003). As mentioned earlier, new millennium learners are not only more skilled and adept at using ICT than their teachers; they have also been shaped by it (Prensky, 2001; Oblinger & Oblinger, 2005; Pedró, 2007) in terms of their patterns of thinking and communication, notions of learning, needs for control, and even their personal and social values. For instance, multitasking is taken for granted by new millennium learners as a normal social practice.

Aside from the different thinking patterns shaped by intensive ICT usage and the various notions of communication and social lifestyles, there are contradictions emerging from contrasting practice in and outside schools (Pedró, 2007). All of these pose challenges to education as well as to ICT integration in the classroom. Research shows that digital students can get more access to ICT at home than in school (Russell, Bebell, & O’Dwyer, 2003; Pedró, 2007). Home access to ICT was found to be substantially related to students’ Internet skills (Kuhlemeier & Hemker, 2007). In addition, a positive conditional relationship with computer use for education and communication at home was found with international student PISA data (Fuchs & Wößmann, 2004). Therefore, the ICT-related behavior outside of institutional settings should be included in the research of digital natives.

As a generation-wide and growing phenomenon (Pedró, 2007), the above general characteristics of digital students have been identified in different investigations around the world. In-depth studies also found that the digital native is by no means an identifiable generation defined solely by age (Bennett & Maton, 2010; Brown & Czerniewicz, 2010); they possess a diverse range of technology skills and preferences (Kennedy et al., 2010; Pedró, 2007; Hofferth, 2010; Li & Ranieri, 2010); their traits are situated and by no means shared across the entire generation (Brown & Czerniewicz, 2010; Sánchez et al., 2011).

The notion of digital native–digital immigrant dichotomy itself will not be addressed in this study, which has gained pervasive discourse without consensus in academic community (e.g., Salajan, Schönwetter, & Cleghorn, 2010; Waycott et al., 2010). Yet the discontinuity that exists between the new millennium students and their teachers is a fact that can be easily observed. This very discontinuity will be the focus of the current investigation, the purpose of which is to guide reconsideration of both the methodology and content appropriate when educating digital natives with ICT integration, both in and outside of institutional settings.

Research framework

This study tries to provide evidence-based understanding about the extent and nature of the barriers to ICT integration from the perspective of technology acceptance, which can be defined as the users’ intention and/or the actual usage of technology. A number of studies have examined the factors that influence user acceptance. In reviewing these studies, we compiled four constructs that have been frequently mentioned as the predictors of ICT acceptance: outcome expectancy, task-technology fit, social influence, and personal factor.
Beliefs on technology use: Outcome

Users’ acceptance of technology was predicted from their internal beliefs and attitudes on their usage, which was measured with perceived usefulness and perceived ease of use in the technology acceptance model (TAM) (Davis, 1989; Venkatesh et al., 2003). It is the usefulness in this model that most strongly predicted the actual use of a technology. In the information system literature, this construct has been addressed as perceived usefulness, relative advantage, performance expectancy, and outcome expectancy in a number of IT acceptance models (Venkatesh et al., 2003), and has been empirically verified as the most important predictor of technology usage (Venkatesh et al., 2003; Kim, Jahng, & Lee, 2007; Lee, 2010; El-gayar, Moran & Hawkes, 2011).

Beliefs on technology use: Task fit

The task-technology fit (TTF) is the degree to which a technology assists an individual in performing his or her tasks (Goodhue & Thompson, 1995; McGill & Hobbs, 2008). This construct has been addressed as effort expectancy in the technology acceptance literature (e.g., El-gayar et al., 2011). Only when the IT application meets the task requirements of users will it have a positive impact on their performance. The assumption of TTF is that users accept technology due to its potential benefits, such as performance improvement, regardless of their attitude. By integrating TTF, Dishaw and Strong (1999) proposed an extended technology acceptance model that is likely to provide a better explanation for IT usage.

Social influence

From the social psychology viewpoint, the dominant social factor is a kind of social norm defined as the “perceived social pressure to perform or not to perform a behavior” (Ajzen, 1991). Lewis, Agarwal and Sambamurthy (2003) suggested and empirically verified that the perceived social influence from referent others has a significant positive influence on individual beliefs about the usefulness of technology. Similarly, recent studies have found that social influence positively and significantly affects IT utilization (Thompson, Compeau & Higgins, 2006; Kim et al., 2007). Extensions to the TAM likewise introduced social norm as an important construct related to beliefs about the usefulness of technology (Venkatesh et al., 2003). The construct of social factors is therefore introduced in our research model to take consideration of both in and out of school ICT usage.

Personal factors

As a study seeks to understand end user technology usage in and out of school, the construct of personal factors is introduced by necessity in the model. Following the interpretation of the individual level factors of technology usage from Lewis et al. (2003), the personal factors include computer self-efficacy and personal innovativeness with technology. Self-efficacy is defined as the belief in one’s capability to perform a particular behavior, it influences decisions about what behaviors to undertake, how much effort it entails, and what emotional responses would be produced (Compeau, Huff & Higgins, 1999). It is widely recognized as one of the explanatory factors that influence end users’ IT usage (Compeau et al., 1999; Lewis et al., 2003; Bandy, Strong & Dishaw, 2006). Personal innovativeness pertains to the degree to which an individual is willing to try out any new information technology (Agarwal & Prasad, 1998). As the most proximate influence on an individual’s cognitive interpretation of information technology, computer self-efficacy and personal innovativeness have been empirically verified as associated with positive technology use (Lewis et al., 2003; Thompson et al., 2006).

Methodology

The target population of the current study is composed of K12 students and teachers in Shanghai. As the most developed metropolis in China, Shanghai has the highest rate of home ICT ownership, with 93% households having computer access whereas 77% having Internet access in 2010 (Li, 2011). Moreover, Shanghai is the leader when it comes to introducing ICT into schools, with 100% of the schools connected to the Internet, and with a computer-student ratio of 3:1 (Cai & Yuan, 2010). The goal of the present study is to investigate Shanghai users’ patterns of
ICT usage, find the difference if it exists, and present suggestions toward a more effective ICT integration in instruction. Data were gathered through surveys, interviews, and school visits. The findings reported in this study are based on the first stage of the questionnaire. The methods of this surveying phase are presented in the following sections: sample and sampling technique, development of instruments, and data analysis.

Sample

Stratified random sample technique was used in this study. First, five districts including two urban and three suburban districts were sampled according to the proportion of all the 19 districts in Shanghai. In each district, five schools were sampled, within which the numbers of primary, middle, and higher schools were counted according to the proportion in the named districts. Ninety students and ten teachers in each school were randomly sampled. Table 1 shows the details of the samples.

![Table 1. Demographic information of samples](image)

<table>
<thead>
<tr>
<th>Type</th>
<th>Grade</th>
<th>Gender</th>
<th>Years of service</th>
<th>Years of ICT use</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher</td>
<td>Primary</td>
<td>Male</td>
<td>≤3 : 28</td>
<td>≤3: 13</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>60</td>
<td>3-5:16</td>
<td>3-5:24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6-9:27</td>
<td>6-9:84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>Female</td>
<td>10-12:26</td>
<td>10-12:75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>89</td>
<td>175</td>
<td>13-15:27</td>
<td>13-15:32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td></td>
<td>16-18:29</td>
<td>16-18:11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td></td>
<td>≥18:96</td>
<td>≥18:9</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>Primary</td>
<td>Male</td>
<td>≤3:608</td>
<td>N/A</td>
<td>2307</td>
</tr>
<tr>
<td></td>
<td>637</td>
<td>1054</td>
<td>3-5:990</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6-9:402</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>Female</td>
<td>10-12:108</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>976</td>
<td>1097</td>
<td>≥13:46</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>548</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instruments

The survey items in the present study were developed by adapting instruments from previous studies. Ten items were developed to measure the outcome expectancy, of which seven were adapted from Venkatesh et al. (2003), McGill and Hobbs (2008) and Thompson et al. (2006); meanwhile, the other three items were tentatively added to measure the expectancy of learning interest and retention. Task-technology fit was measured with five items adapted from McGill and Hobbs (2008), covering the aspects of work compatibility, ease of use, and information quality. Social influence was measured with five items adapted from McGill and Hobbs (2008) and Thompson et al. (2006). The items referred to the participants’ beliefs on whether the organization or other individuals want them to use the information technology. Personal factor was measured with seven items from two aspects: four items of computer self-efficacy were adapted from Venkatesh et al. (2003), whereas the other three items of personal innovativeness with ICT were from Thompson et al. (2006). The wording of questions for the teacher version and the student version was adjusted to suit each context. Except for the construct of individual factors, the items of the other three constructs were measured both for inside and outside school usage. All of the items were measured using a 7-point Likert scale ranging from “strongly agree” to “strongly disagree.” A sample of items from these scales can be found in Table 2.

The responses to the survey were subjected to factor analysis to verify the constructs for inside and outside use of both teachers and students. The Statistical Package for the Social Sciences (SPSS V.16) and its principle components analysis method was used to extract the relevant number of factors, and Varimaxrotation was used to achieve simple structure. The factors with salient loadings were subjected to verify the constructs in the research model. The constructs in the research model were updated accordingly, and the final scores on each construct were calculated with the mean of its constituent items. Table 2 shows the questionnaire items for students about their classroom ICT usage and their factor loadings on the constructs. The reliability was estimated using Cronbach’s coefficient alpha.
Table 2. Items and their factor loading for students in classroom ICT usage

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OE(α = .931)</td>
</tr>
<tr>
<td>Using ICT in class increases my learning efficiency.</td>
<td>.793</td>
</tr>
<tr>
<td>Using ICT in class enables me to study better.</td>
<td>.784</td>
</tr>
<tr>
<td>I find the ICT in the classroom useful in my learning.</td>
<td>.738</td>
</tr>
<tr>
<td>If I use ICT in class, I can learn better and more quickly.</td>
<td>.728</td>
</tr>
<tr>
<td>Using ICT improves the quality of my learning in class (eg. more focused)</td>
<td>.723</td>
</tr>
<tr>
<td>Using ICT can improve my interest in learning.</td>
<td>.699</td>
</tr>
<tr>
<td>Using ICT can help me keep a longer focused learning state.</td>
<td>.649</td>
</tr>
<tr>
<td>Using ICT can help me remember the learning content more firmly.</td>
<td>.599</td>
</tr>
<tr>
<td>I can complete the task using ICT even if no one around to tell me what to do.</td>
<td>.187</td>
</tr>
<tr>
<td>I can complete the task using ICT even only with the built-in help (such as a</td>
<td>.205</td>
</tr>
<tr>
<td>tutorial).</td>
<td></td>
</tr>
<tr>
<td>I can complete the task using ICT if I had enough time with the technology</td>
<td>.190</td>
</tr>
<tr>
<td>provided.</td>
<td></td>
</tr>
<tr>
<td>I can complete the task using ICT if I can get help when I am stuck.</td>
<td>.245</td>
</tr>
<tr>
<td>My classmates use a lot of ICT in or out of class.</td>
<td>.189</td>
</tr>
<tr>
<td>I would use ICT for fun or for other needs outside the classroom.</td>
<td>.187</td>
</tr>
<tr>
<td>I am in the first group to use ICT in learning.</td>
<td>.156</td>
</tr>
<tr>
<td>ICT is compatible with all aspects of my tasks.</td>
<td>.315</td>
</tr>
<tr>
<td>It is easy to get ICT to do what I want it to do.</td>
<td>.194</td>
</tr>
<tr>
<td>ICT fits well with the way I like to work.</td>
<td>.307</td>
</tr>
<tr>
<td>I can access ICT whenever I need it.</td>
<td>.189</td>
</tr>
<tr>
<td>ICT in class is easy to use.</td>
<td>.360</td>
</tr>
<tr>
<td>Using ICT inside the classroom contributes to the interaction between my teacher</td>
<td>.499</td>
</tr>
<tr>
<td>and I/my classmates and I.</td>
<td></td>
</tr>
<tr>
<td>In-class use of ICT makes it easier to understand the teachers.</td>
<td>.480</td>
</tr>
<tr>
<td>My teachers expect me to use ICT in the classroom.</td>
<td>.264</td>
</tr>
<tr>
<td>My parents expect me to use ICT in the classroom.</td>
<td>.269</td>
</tr>
<tr>
<td>I seldom try new technology in my study before I see other students using it.</td>
<td>-.102</td>
</tr>
<tr>
<td>My school expects me to use ICT in the classroom.</td>
<td>.291</td>
</tr>
<tr>
<td>People whose opinions I value and trust expect me to use ICT in my learning.</td>
<td>.261</td>
</tr>
</tbody>
</table>

Note. OE: outcome expectancy; PF: personal factors; TTF: task-technology fit; SF: social factors

In addition, ICT use duration and frequency, in-class and outside of school, for both teachers and students were measured to approximate the actual degree of IT use. The current study did not target a particular technology. Rather, ICT was used as a generic term to cover all kinds of contemporary technologies. This definition of technology was provided in the questionnaire. The score was obtained by employing a Likert-type scale ranging from “never” to “very frequently.” The duration of technology usage for inside and outside school was measured using a Likert-type scale time ranging from “three years or less” to “19 years or more.”

Data analysis and results

Students’ ICT usage

The list of contemporary technologies students are using is quite long. Except for standard office package applications and the typical educational software such as Geometer’s Sketchpad, students have much more opportunities to use various technologies outside the school than in the classroom. Detailed percentile breakdowns of the most frequently used technologies in two settings could be found in Figure 1.

In general, students have a longer experience in using ICT at home than in school. Almost 70% of them have been using ICT for three years or more, whereas their classroom use is approximately 40%. On the other hand, when
comparing the frequency and time of ICT usage, students use ICT more in the classroom than outside the school (M = 2.52, SD = 0.68 and M = 1.94, SD = 0.72 for in and outside use). Pairwise T-test showed that this difference is significant (t = 2.087, p = .000).

Within the three grade levels, significant differences in ICT usage were found both in and outside the school. Post hoc of ANOVA test revealed that pairwise differences existed among all the three grades. In terms of in-class ICT usage, high school students had a higher in-class ICT usage than middle school (mean difference = .243, p = .000) and primary school students (mean difference = .386, p = .000); on the other hand, middle school students had a higher in-class ICT usage than primary school students (mean difference = .386, p = .000). The same pattern of differences emerged in terms of outside-school ICT usage. The mean differences between high school students and middle, primary, and between junior and primary school students were 0.477 (p = .000), 0.205 (p = .000), and 0.272 (p = .000), respectively. These differences were confirmed when the ANOVA test was conducted within these age groups. Differences were also found between girls and boys. For in-class ICT usage, that of girls was significantly higher than the boys’ (p = .000); meanwhile, for outside-school ICT usage, that of boys was significantly higher than the girls’ (p = .000).

As can be seen in Table 2, the questionnaire items load onto four factors that are presumably related to students’ use of ICT both in and outside the classroom. This could preliminarily verify the four constructs in the research model. However, there are two items originally developed to measure outcome expectancy on the task-technology fit for in-class use, and one item originally developed to measure outcome expectancy on task-technology fit.

The mean scores for each of the four factors were calculated, and the scores both for inside and outside classroom usage were subjected to the pairwise T-test. The results showed significant differences in student perception toward the usefulness of ICT inside and outside the classroom (t = 3.917, P = .000). Obviously, students had higher expectancy for in-class ICT usage (M = 5.589, SD = 0.991) than outside school (M = 5.5286, SD = 0.999). Social influence that affects ICT usage was also significantly different between inside and outside classroom usage (t = 3.466, p = .001); outside usage was affected more by social influence (M = 5.344, SD = 1.073) than that of in-class usage (M = 5.304, SD = 1.08). Task-technology fit did not exhibit any difference between inside and outside ICT usage.

Within the different grades, the ANOVA test showed a significant difference in outcome expectancy for in-class ICT usage (F = 10.79, p = .000). Both middle and primary school students held higher expectancy for ICT in classrooms than their counterparts in high school (mean difference = 0.175, p = .003; mean difference = 0.263, p = .000, respectively). Both middle and primary school students perceived a better task-technology fit than their counterparts in high school, both for in-class and outside ICT usage. Social influence for in-class ICT usage exhibited significant differences between middle and high school students. Moreover, significant differences only existed because of personal factors. In general, no significant differences were found between primary and middle school students. In addition, boys were influenced socially for their ICT usage both in and outside the school more than girls (p = 0.021 and 0.005, respectively); however, no differences in outcome expectancy, perception of task-technology fit, and personal factors existed between the gender groups.

Teachers’ ICT usage

The technology that teachers most frequently used in classrooms consists of office programs, such as Microsoft Office and WPS. Teachers also used multimedia heavily. Video was used most frequently (67.07%) compared to other media programs (30.52%). Except for technologies of class preparation, the next most frequently used technologies outside class are those for social connection. Details of the percentile breakdowns of the types of technologies that teachers use could be found in Figure 1.

In general, teachers used ICT more in classrooms than outside of school (M = 3.40, SD = 0.77 and M = 2.95, SD = 0.87 for in and outside use). The pairwise T-test showed that this difference is significant (t = 9.345, p = .000). The grade levels that teachers taught did not exhibit any difference for ICT usage in or outside the school, which was discrepant from the perception of students especially for in-class ICT usage. No gender difference toward ICT usage was found. Differences in in-class ICT usage were found within the groups of teachers with different service years. Teachers who have served less than five years used ICT less in the classrooms compared to their counterparts from
all the longer service year groups. With regard to outside-school usage, no significant difference was exhibited between the different service groups.

A similar factor analysis was carried out for the teachers’ questionnaire. The same pattern turned out, with four factors extracted as important to teachers’ ICT usage outside the school, which could also preliminarily verify the four constructs in the research model. However, three factors were extracted when the same factor analysis was conducted for the data of in-class ICT usage, where the items of task-technology fit and those of personal factors were loaded on the same construct. These constructs were interpreted as personal factors in such a way that the perception of task-technology fit may be taken for teachers’ self-efficiency and competence of ICT in classroom integration.

The mean scores for each of the factors were calculated, and the scores both for inside and outside classroom use were subjected to the pairwise T-test where it was feasible. Unlike the perception of students, teachers had a significantly higher expectancy for the usefulness of ICT outside of class than what was integrated into their instruction ($t = 7.244, p = .000$). The comparison of social influences affecting their inside and outside school ICT usage was likewise significantly different. Apparently, teachers’ in-class ICT integration was affected more by social influence ($M = 5.117, SD = 0.726$) than that of outside school usage ($M = 5.028, SD = 0.759$).

**Comparison of ICT usage between students and teachers**

The differences in the frequently used ICT between students and teachers can be found in Figure 1. In general, the variety of frequently used technologies in-class was much less than that of the outside, office software and media players has almost dominated. Students have many more ICT choices outside the school than in the classroom. In terms of the types of ICT used, students used more types than teachers did.

![Figure 1. The percentile breakdowns of used technologies](image)

With regard to the difference in duration and frequency of ICT usage between student and teacher groups, the independent T-test revealed that whether ICT was used in or outside the classroom did not matter; teachers’ use exceeded that of students ($t = 20.01, p = .000$ and $t = 17.01, p = .000$ respectively). The different expectations of teachers and students when it comes to ICT in-class and outside were mentioned previously. When comparing the expectations between the two groups, teachers had significantly higher perceptions on the usefulness of ICT outside of class ($M = 5.80$ versus $M = 4.89$, $t = 16.995, p = .000$), whereas there was no significant difference for in-class
ICT integration. Likewise, the different impacts of social influence on ICT use in and out of classrooms were earlier reported. When comparing the influence of social factors between the two groups, teachers were affected more than students by social influence, both for in-class integration and outside of class usage ($t = 5.147, p = .000$ and $t = 4.005, p = .000$).

In terms of the difference of personal factors between students and teachers, results showed that students had higher self-efficiency and confidence in using ICT than teachers ($M = 5.59$, SD = 0.97 versus $M = 5.25$, SD = 0.96; $t = 5.209, p = .000$). Factor analysis preliminarily tested the constructs related to ICT usage in the research model, while those factors have different effects on in-class integration and outside school usage for teachers and students. Knowing how these factors could predict overall ICT usage as well as the ICT integration into class and outside the school would be informative. Linear regression was used to determine how those factors could explain usage. Data for students and teachers were collapsed to conduct a linear regression test for overall ICT usage in and outside of class. The results showed that the most powerful predictor is the personal factor, getting beta = 0.201 ($t = 6.371, p = .000$) for in-class ICT usage and beta = 0.217 ($t = 6.552, p = .000$). Although task-technology fit yielded a similar coefficient for in-class and outside ICT predicting (beta = 0.188, $t = 4.759, p = .000$ and beta = 0.186, $t = 5.262, p = .000$, respectively), outcome expectancy and social influence obtained radically discrepant results: for in-class situations, social influence explained ICT usage significantly (beta = .082, $t = 2.843, p = .005$) but not for outside situations. Outcome expectancy explained ICT usage outside the school significantly (beta = .070, $t = 2.190, p = .029$) but not for in-class use. These results were echoed in separate regressions with teacher data and student data, where the personal factor is the most powerful predictor; outcome expectancy can only predict the outside ICT use, and social factors only predict the in-class ICT usage.

**Discussion and conclusion**

Keeping in mind the fast growth of the digital-native phenomenon in the student population, the current study attempted to determine whether the ICT used in-class designed by digital immigrants for the benefit of digital natives meets the expectations of students and complies with their established ICT practice. The findings reported in this study focused on the current situation of the end users' acceptance of ICT both used in-class and outside the school, the differences in the influencing factors, namely, their attitudes, the self-efficiency and competency toward ICT, and how the ICT usage of student and teacher groups is affected by those factors. The discussion of the findings is as follows.

First, the findings of this investigation present evidence of the characteristics of digital natives in Shanghai, both in terms of their perceptions of ICT currently integrated into classrooms and outside the school, as well as the differences between them. As a growing and generation-wide phenomenon (Pedrò, 2007), digital students have been scrutinized in previous studies on subjects such as their general characteristics (Pedrò, 2007), their different types of ICT usage (Kennedy et al., 2010), and their competency in ICT skills (Li & Ranieri, 2010). The findings of this investigation confirm that the digital native is a complicated phenomenon (e.g., Salajan et al., 2010; Kennedy et al., 2010), indicating that in terms of the duration and frequency of ICT usage, students do not use more ICT than teachers do; in terms of the years of ICT usage and the types of ICT being used, students exhibit the typical “native” characteristics of high percentage of early adoption of ICT. Our findings also echo the claim that students can get more ICT access at home than in school (Russell et al., 2003; Pedrò, 2007) in terms of the variety ICT used.

The findings of this investigation present a complex pattern of differences of ICT usage between different groups and different settings. These differences need to be explained by a combination of factors, such as classroom control, and learning tasks that need ICT. The findings show that students are not using as much ICT as their “immigrant” teachers, and that they are not using as much ICT at home as in class. Meanwhile, within the different grade groups, the grade level is positively correlated with reported ICT usage. It is easy to attribute the low usage in the classroom to the teacher’s control; on the other hand, it could be quite complicated to account for the low usage outside the school when home ICT access is not an issue for most students. The most plausible reason could be that there are only a few learning tasks assigned to students to use ICT at home. Further information is needed to verify if the higher grade teachers might be assigning more learning tasks that need ICT access and lead to the higher reported home ICT usage for higher level students than that of the lower levels.
The present study investigated the factors that presumably affect ICT usage for end users, as well as the possible differences between these groups. The four constructs have been verified using the factor analysis approach, except for the teachers’ in-class ICT integration. For the ICT usage in different situations, however, these constructs have revealed different perceptions of two user groups.

In general, personal factors are most important for either students or teachers, both in and outside the school. The significantly higher self-perceptions on ICT of students than that of teachers fit the general description of digital students, although teachers in Shanghai have been undergoing training to develop their competency in ICT integration for years (Zhang & Zhang, 2009). In addition, from the findings of the factor loading with teachers’ data, we can interpret that teachers perceive technology as a way to meet their tasks as part of the ICT competency. In their perception, self-efficiency and ICT competency are the most important factors in determining their ICT adoption in classroom teaching. This finding echoes other studies focusing on the barriers to ICT integration in the classroom, which frequently find that the confidence, competence, and attitudes of teachers are most vital for successful ICT integration (Hew & Brush, 2007; Bingimlas, 2009).

Similar to the original technology acceptance model (Davis, 1989; Venkatesh et al., 2003) and later related research, the factor of outcome expectancy has been verified as one of the constructs related to ICT usage, which has significantly explained the ICT usage outside the school. By comparing the outcome expectancy between different groups and settings, we learn that students hold higher expectancies for in-class ICT integration than their teachers. The younger the students, the higher their expectancies tend to be. Thus far, we have not yet collected the data on how teachers are in generally integrating ICT into different grades, except those on the most frequently used technologies. Therefore, a tentative interpretation could be that the way in which ICT is being integrated is not complying with “native” students who have already established their ICT practices and habits and therefore their expectations. The fact that students are using much more technologies outside the class than in school may confirm this interpretation. The lower expectations of teachers for in-class ICT integration than the use outside the school confirm that in-class integration of ICT is not fulfilling its potential.

Unlike the positive and significant social influence of ICT usage found in previous research (e.g., Lewis et al., 2003; Thompson et al., 2006), the current study finds that social influence has a radical impact on ICT adoption for in-class integration and outside use for both groups. Inconsistent with the stereotype of digital children who tend to follow their peers in adopting new technologies, although the mean score of social influence is quite high (5.344 and 5.304), we find that social influence does not significantly predict the use of ICT outside of school; rather, it is significant for in-class usage. The same pattern has been found for teachers.

With regard to the most powerful predictor of ICT adoption of users (Davis, 1989; Venkatesh et al., 2003), the findings in the current study are discrepant: social influence together with the teachers’ personal factors instead of the outcome expectancy is the significant predictors for in-class ICT integration. This finding deserves further empirical investigation. Although personal factor is readily interpreted as an important determining factor, the role that social influence has on the teachers’ adoption of ICT could be interpreted as a kind of “social pressure” that teachers are perceiving in an educational situation where ICT investment keeps growing, and the students’ ICT knowledge and skills keep developing.

Although the task-technology fit model assumes that users accept technology due to its potential benefits regardless of user attitude (Dishaw & Strong, 1999), the findings of the present study do not support this idea. Task-technology fit is most likely perceived by teachers as a part of ICT self-efficiency, in a sense that as their ICT competency developing, their perception of task-technology fit increases as well. This has been found in the factor analysis with the teachers’ data. Digital children’s perception of task-technology fit is generally high (M = 5.60 and 5.63) across age and gender groups.

As the first phase of a study that attempts to determine the end-user difference of ICT usage between students and teachers and the related factors, we find that this study shows value in compiling a research model and using that model to compare the end users’ behavior and attitudes toward ICT. The results that present the current situations of ICT usage both integrated into classrooms and used outside of schools should also contribute to research toward a better understanding of new millennium learners and to better educate the digital natives with ICT integration. However, there are limitations in this study: First, bias is inevitable due to self-reporting of the subjects, further information is needed to verify the reporting results; second, situated nature of the results is unavoidable, it should be
considered in making sense of the current situations of ICT usage. The next step of this study is to gather further data through interviews and school visits to make in-depth interpretations toward some of the preliminary findings presented at this point.

Acknowledgments

This study is supported by Program for New Century Excellent Talents in University (NCET). We would like to acknowledge the teachers and students who participated in this study at the five districts in Shanghai. We are also very grateful to the anonymous reviewers for his/her insightful and knowledgeable comments and constructive suggestions.

References


Effects of Video Caption Modes on English Listening Comprehension and Vocabulary Acquisition Using Handheld Devices

Ching-Kun Hsu¹, Gwo-Jen Hwang²*, Yu-Tzu Chang³ and Chih-Kai Chang³

¹Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan // ²Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // ³Department of Information and Learning Technology, National University of Tainan, Taiwan // chingkun.hsu@gmail.com // gjhwang.academic@gmail.com // t09955018@stumail.nutn.edu.tw // chihkai@mail.nutn.edu.tw

*Corresponding author

(Submitted October 15, 2011; Revised February 19, 2012; Accepted February 22, 2012)

ABSTRACT
This study investigates the effects of different display modes of video captions on mobile devices, including non-caption, full-caption, and target-word modes, on the English comprehension and vocabulary acquisition of fifth graders. During the one-month experiment, the status of the students' English listening comprehension and vocabulary acquisition was evaluated on a weekly basis. From the experimental results, it was found that the learning achievement of the English target-word group was as good as that of the full-caption group in terms of vocabulary acquisition, while both groups outperformed the non-caption group. When the students' learning preferences were taken into consideration, the visual style students in both the English target-word group and the full-caption group showed significantly better learning effectiveness in terms of vocabulary acquisition than those in the control group. Furthermore, in terms of listening comprehension, the students in all three groups made remarkable progress.

Keywords
Listening comprehension, Vocabulary acquisition, Captions, Mobile learning, Learning style

Introduction
English has been recognized as being an important language for international communications. In the past decades, many non-English speaking countries have developed and utilized various computer systems to support English as Foreign Language (EFL) learners. Moreover, most of these countries intend to have their children learn English as early as possible; for example, the Ministry of Education in Taiwan has extended regular English instruction to the third grade of elementary school and is planning to initiate an English learning program for first graders within ten years.

Foreign language learning can be considered from the four aspects of listening, speaking, reading, and writing. Among these aspects, listening is an important capability of social interactions, and it has been found that people receive new messages more efficiently via listening than reading (Luo, 2008). The Ministry of Education in Taiwan even guides elementary schools to put emphasis on English listening first, followed by reading or writing. In the meantime, most forms of English certification include listening proficiency tests, implying the importance of fostering English listening competences.

With the advance of mobile technologies and multimedia, instructional materials for English listening training have been developed in a variety of forms. For example, MP3 players have become a new mobile device for learning via listening; moreover, several web2.0 media, such as YouTube, have become popular channels that provide audio and video materials for language learning (Godwin-Jones, 2007). In the meantime, the popularity of various mobile devices (e.g., smartphones and e-books) and wireless networks (e.g., Wi-Fi and Wi-Max) have enabled individual students to use those emerging learning materials or channels anywhere and at any time (Hwang, Shi, & Chu, 2011; Hwang, Wu, Tseng, & Huang, 2011; Wu et al., 2011). Therefore, it can be foreseen that students will eventually be equipped with a mobile device installed with proper learning tools, systems or materials so that they can make their own learning progress, and even learn with content of appropriate difficulty to match their level of proficiency (Hung et al., 2012; Hwang & Chang, 2011; Hwang, Chu, Lin, & Tsai, 2011; Norris, Hossain, & Soloway, 2011). In recent years, several studies concerning the use of mobile technology in language learning have revealed the benefits of this approach, such as the provision of opportunities for individual practice and a seamless learning environment (Ogata,
Matsuka, El-Bishouty, & Yano, 2009; Ogata & Yano, 2004; Ozcelik & Acarturk, 2011; Wong, Chin, Tan, & Liu, 2010).

On the other hand, previous studies have reported that videos embedded with captions are helpful for students in learning second language reading (Chun & Plass, 1997) and listening (Danan, 1992). Hsu and Chang (2010) have further indicated that hiding the easier vocabulary and showing only the relatively difficult words in the captions may contribute to undergraduates' listening comprehension. Consequently, this study developed different display modes of captions on mobile devices for improving English listening competence and promoting the vocabulary acquisition of students. The study aims at exploring whether different display modes of captions and subtitles result in different degrees of effectiveness in the listening comprehension and vocabulary acquisition of elementary school students. Moreover, the learning performance of students with different learner preferences is also compared. Furthermore, the students' perceived satisfaction with the approach, and their perceptions of the usefulness and ease-of-use of the mobile learning activity are reported as well.

Literature review

Video captions and English listening comprehension

Using videos or films as a learning resource has received a great deal of attention from researchers and has been successfully applied to various educational applications (Yang, Huang, Tsai, Chung, & Wu, 2009). Researchers have indicated that multimedia learning materials are more useful than traditional paper-based instruction (Mackey & Ho, 2008; Mayer & Moreno, 2002; Rose, 2003). Videos which provide visual, contextual, and non-verbal input provide foreign language learners with simultaneous visual and aural stimuli which can make up for any lack of comprehension resulting from listening alone (Brett, 1995; Hoven, 1999; Seo, 2002). Several previous studies have shown that such videos are highly accepted by learners during the learning process (Choi & Johnson, 2005; Choi & Johnson, 2007; Mackey & Ho, 2008). Moreover, captions have been perceived as the most useful and efficient auxiliary for watching videos (Hsu, 1994), while videos with authentic accents have been recognized as being a good learning resource for language learners (Dahbi, 2004).

Researchers have indicated that students need to receive a great quantity of comprehensible input so as to achieve the objective of language learning when they learn a foreign or second language (Krashen, 1981; 1985). When students watch videos in a foreign language, the contribution of comprehending and connecting the foreign language and its meaning is limited if they cannot understand what they have heard. Therefore, using captions and subtitles is useful in enhancing the effectiveness of listening comprehension. Scholars have defined subtitles as the on-screen text in the students’ native language combined with a second language soundtrack in the video, while captions are the on-screen text in the original language combined with a soundtrack in the same language (Markham, Peter, & McCarthy, 2001). Therefore, in this study, the term subtitles refers to on-screen Chinese text combined with an English soundtrack, while captions refers to on-screen English text combined with an English soundtrack. In addition, bilingual subtitling has been defined as on-screen text in both the students’ native language and the target language combined with a target language soundtrack (Katchen et al., 2001). That is, bilingual subtitling as used in this study refers to English audio with the simultaneous appearance of English and target-word Chinese texts on the screen. Furthermore, the target words represent the new or key vocabulary of the target lessons in the following discussions.

Captions and subtitles were initially used in foreign language instruction in the 1980s, and many studies have confirmed that combining captions with audio-visual materials is an effective instructional method to enhance the listening and reading comprehension of a second language (Borras & Lafayette, 1994; Danan, 2004; Garza, 1991; Markham & Peter, 2003) because learners can confirm the information they hear by way of the support provided by the captions (Froehlich, 1988; Grimmer, 1992; Vanderplank, 1988). In other words, captions visualize the auditory information of the foreign language which the learners hear in the video (Danan, 2004). Because of such additional cognitive processing, captions and subtitles promote language comprehension (Bird & Williams, 2002). Researchers have further indicated that videos with captions facilitate vocabulary acquisition (Plass, Chun, Mayer, & Leutner, 1998), reading (Chun & Plass, 1997), and listening comprehension (Danan, 1992; Hsu, 1994; Markham & McCarthy, 2001).

On the other hand, Vandergrift (2007) stated that providing native subtitles for learners will obstruct their familiarity with pronunciation. Among many different assisted methods for learners who use videos to train their foreign
language listening, Liou (2000) found that the highest proportion of students used foreign language captions as auxiliary support, especially the higher-achieving students. The replay function was the next, and the assistance of native subtitles was the third ranked strategy when learners needed support. Therefore, adopting an innovative approach, this study provided either full or partial foreign language captions, but did not provide full native subtitles for the learners in any of the three groups.

Mobile assisted language learning

Mobile learning creates diverse possibilities for innovative instructional methods to be carried out in the general classroom in more effective and efficient ways. Scholars have noted that mobile-assisted language learning (MALL) provides students with rich, real-time, convenient, and contextual learning opportunities, no matter whether they are inside or outside the classroom (Kukulska-Hulme, & Shield, 2008). At the present time, MALL is not only one of the main language learning resources for students, but it also contributes to the learning and utilization of new language skills (Hashemi, & Ghasemi, 2011). Scholars have found that learners pay most attention to captions, followed by video and audio, and acquire most words by associating them with visual images. They have therefore concluded that captioned video tends to aid recognition of written word forms and the learning of word meaning. On the other hand, non-captioned video tends to improve listening comprehension as it facilitates recognition of aural word forms (Winke, Gass, & Sydorenko, 2010).

Recently, an increasing number of MALL studies have been devoted to vocabulary acquisition and learning, and the learning outcomes have been significant (Chang & Hsu, 2011; Chen, & Chung, 2008; Lu, 2008; Kim, & Kim, in press). In one study, students were encouraged to use mobile phones to access video clips explaining English idioms (Thornton & Houser, 2005). This is part of an obvious trend that learners are spending more time using mobile devices rather than PCs to carry out learning activities (Stockwell, 2010). In another study, the language learners expressed positive attitudes toward using mobile phones to browse wireless application protocol sites, which was also found to be effective for developing listening skills and for student-centered learning (Nah, White, & Sussex, 2008). In virtue of the benefits of MALL, and as many students nowadays are so familiar with mobile devices and are used to applying them in learning, this study conducted a number of MALL instructional experiments, in which different modes of captions and subtitles were inset in the MALL video material for different groups, with the purpose of identifying better instruments for English as Foreign Language learners (i.e., EFL learners).

Learning style and the Technology Acceptance Model

Learning style refers to individual preferences regarding ways of learning, and affects how individuals accept stimuli, memories, thinking, and problem-solving. There are many different scholars who have proposed diverse categories of learning styles (Reichmann & Grasha, 1974; Fischer & Fischer, 1979; Keefe, 1979a; 1979b; Gregorc, 1979; Schmeck, 1982; Kolb, 1984; Myers & McCaulley, 1985; Honey & Mumford, 1986; Felder & Silverman, 1988). If teachers realize the difference in learning styles among learners, and design appropriate instructional methods or media, learners are likely to benefit. Taking the different learning styles of visual and verbal learners for example, visual learners remember best what they see, such as pictures, diagrams, flow charts, time lines, films, and demonstrations. Verbal learners on the other hand, get more out of words, such as written and spoken explanations. As this distinction between the visual/verbal dimensions is highly relevant to the use of videos in training the listening competence or vocabulary acquisition of a foreign language, this study employed the visual/verbal learning style dimensions to evaluate the learning preferences of the participants.

TAM is a well-known model proposed by Davis (1986) based on the Theory of Reasoned Action (TRA). This model provides a theoretical foundation for understanding how external variables influence the inner beliefs, attitudes and intentions of users, and then affect their use of technology. It also provides a way to explain users’ behavior when accepting new information technology (IT) and to analyze the factors that influence their attitudes toward its use (Davis, 1989; Venkatesh & Davis, 1996). Perceived usefulness (PU) refers to the degree to which people believe that using a particular system would enhance their performance, while perceived ease-of-use (PEOU) refers to the degree to which people believe that using a particular system would be free from effort (Davis, 1989). This study uses the TAM questionnaire to investigate learners' perceptions of the different caption and subtitle modes inset in the MALL materials (i.e., video).
Method

Participants

The experiments were conducted in an elementary school in Taiwan, where the students learn English as a foreign language. The school categorizes the English proficiency of the students into three levels, A, B, and C, based on an English proficiency test developed by a group of experienced English teachers. The participants of this study were three classes of fifth graders who were 11 years old on average and were categorized as having an English proficiency of level C; that is, they were low-achievement students in English. During the learning activity, one class of 27 students, including 16 males and 11 females, was assigned to be control group one, another class of 28 students, including 12 males and 16 females, was assigned to be experimental group one, and the other class of 26 students, including 15 males and 11 females, was experimental group two.

Research design

The experiment was conducted for a month, as shown in Figure 1. The participants used PDAs (Personal Digital Assistants) to play the instructional video related to the lesson they studied in class each week. After spending time watching the video, they immediately took a test to evaluate their listening comprehension proficiency and vocabulary acquisition. Although this study provided PDAs as dedicated devices, it should be noted that other mobile devices, such as smartphones or e-books, could also be used to conduct the same learning activity.

During the learning activity, a video without any English captions or Chinese subtitles was used for the students in the control group because previous studies have indicated that providing no captions or subtitles helps students adapt to various pronunciation features, such as reduced forms, assimilation, elision, and re-syllabification (Vandergrift, 2007).

On the other hand, a video with full English captions and Chinese subtitles of only the target vocabulary was used for the students in experimental group one because a previous study showed that such a setting is helpful in training listening proficiency and comprehension, and confirmed that full Chinese subtitles are not needed (Hsu & Chang, 2010). Another video with both English captions of the target vocabulary and Chinese subtitles of the target vocabulary was used for the students in experimental group two.

To summarize, the videos for the three groups consisted of the same content with different caption modes. No captions were provided for the control group, while full English captions and Chinese target words were provided for
experimental group one, and English and Chinese target words were provided for experimental group two. Figure 2 shows an example of the caption modes for experimental group one (left) and experimental group two (right).

Figure 2. The playing interfaces of experimental group one (left) and two (right)

Researchers have indicated that, for students to get used to the tempo of usual conversation, there is no need to provide fast forward or slow play functions; instead, only the functions of play, pause, and replay are necessary for listening training (Grgurovi & Hegelheimer, 2007; Winke, Gass, & Sydorenko, 2010). In order to take part in the mobile assisted listening training, the students in each group could use a stylus to select the functions of play, pause, and replay to listen in the limited time, as shown in Figure 3. As shown in the left photo of Figure 3, a timer was used to remind the students of the time limitation while watching the video. After conducting the post-test and administering the TAM questionnaire, this study further interviewed one third of the students (i.e., nine students from each group) to collect more detailed and in-depth feedback.

Figure 3. The students use PDAs to facilitate their English listening training

Research tool

The research tools in this study included the visual/verbal learning style measure, learning achievement tests, and the questionnaire for measuring the students’ technology acceptance.

The test sheets were developed by two experienced teachers. The pre-test consisted of two groups of questions about the students’ prior knowledge of the listening comprehension content and their vocabulary capability. It consisted of twenty-six multiple-choice items for examining listening comprehension and twenty-four multiple-choice items for testing vocabulary capability, with a perfect score of 100, and was also the school’s midterm English exam. The weekly assessment and the post-test all consisted of five multiple-choice items for evaluating the students’ listening comprehension, and five multiple-choice items for assessing the vocabulary acquisition of the lesson unit. The perfect score of the weekly tests and the post-test was 100. Each listening comprehension test was audio broadcast, and the students were asked to listen to the questions and fill out the answers on the answer sheet.
The Index of Learning Styles (ILS) Questionnaire was developed by Felder and Soloman (1991) based on the learning styles proposed by Felder and Silverman (1988). The original ILS measure consists of four dimensions (i.e., sensing/intuitive, visual/verbal, active/reflective and sequential/global), each of which contains 11 items. In this study, the "visual/verbal" dimension was adopted. The visual/verbal dimension consists of 11 items with a Cronbach’s alpha value of 0.74 (Litzinger, Lee, & Wise, 2005).

In addition, this study employed the Technology Acceptance Model (TAM) questionnaire recently revised from the TAM questionnaire for mobile learning (Chu, Hwang, & Tsai, 2010; Chu, Hwang, Tsai, & Tseng, 2010). Three dimensions are explored in the questionnaire, that is, ease-of-use, perceived usefulness, and perceived satisfaction. The questionnaire uses a six-point Likert-scale, where 1 represents “strongly disagree” and 6 represents “strongly agree.” The Cronbach’s alpha values for the three dimensions, ease-of-use, perceived usefulness, and perceived satisfaction, are 0.88, 0.98 and 0.91, respectively.

Experimental results

Analysis of the pre-test and post-test

One of the objectives of this study was to examine the effectiveness of the different caption and subtitle modes with respect to the students’ listening comprehension and vocabulary acquisition. It was therefore first necessary to ensure that the three groups had comparable listening comprehension and vocabulary levels before beginning the treatment. This was determined by way of the pre-test for which the mean values and standard deviations of the scores were 42.07 and 5.07 for the control group, 41.57 and 5.80 for experimental group one, and 42.15 and 8.27 for experimental group two. The ANOVA analysis results of the pre-test among the three groups do not show a significant difference ($p = 0.94 > .05$); that is, it was ascertained that the three groups of students had equivalent prior knowledge before the learning activity.

In addition, by employing ANCOVA on the post-test scores of the three groups, no significant difference was found between the listening comprehension scores of the three groups ($F = 1.94, p > .05$), while significant differences were found between the vocabulary acquisition scores of the experimental groups and those of the control group ($F = 3.71, p < .05$), as shown in Table 1. As the adjusted means of experimental group one (68.81) and experimental group two (67.17) were both significantly higher than that of the control group (51.10), the students who watched the videos with captions (no matter whether they were full English captions with Chinese target words or only English target words with Chinese target words) revealed significantly better learning achievements in English vocabulary than those who learned without captions. Therefore, the provision of Chinese target-word subtitles could be the crucial factor in promoting the effectiveness of vocabulary acquisition.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted Mean</th>
<th>F</th>
<th>Pairwise comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group (a)</td>
<td>27</td>
<td>48.15</td>
<td>24.34</td>
<td>51.10</td>
<td>3.71*</td>
<td>(b) &gt; (a)*</td>
</tr>
<tr>
<td>Experimental group one (b)</td>
<td>28</td>
<td>71.43</td>
<td>26.35</td>
<td>68.81</td>
<td></td>
<td>(c) &gt; (a)*</td>
</tr>
<tr>
<td>Experimental group two (c)</td>
<td>26</td>
<td>70.00</td>
<td>33.11</td>
<td>67.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*$p < .05$

Table 1. ANCOVA analysis results of the vocabulary acquisition tests

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre-test Mean</th>
<th>Post-test Mean</th>
<th>t</th>
<th>Pre-test Mean</th>
<th>Post-test Mean</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group (a)</td>
<td>27</td>
<td>42.07</td>
<td>65.93</td>
<td>-5.21**</td>
<td>23.90</td>
<td>48.15</td>
<td>-5.47**</td>
</tr>
<tr>
<td>Experimental group one (b)</td>
<td>28</td>
<td>41.57</td>
<td>77.14</td>
<td>-8.24**</td>
<td>26.61</td>
<td>71.43</td>
<td>-9.70**</td>
</tr>
<tr>
<td>Experimental group two (c)</td>
<td>26</td>
<td>42.15</td>
<td>71.54</td>
<td>-6.81**</td>
<td>26.56</td>
<td>70.00</td>
<td>-7.24**</td>
</tr>
</tbody>
</table>

*$p < 0.01$

Table 2. Paired sample t-test between the pre-test and the post-test

Table 2 shows the paired sample t-test results between the pre-test and the post-test. It can be seen that the three groups all made significant progress in listening comprehension and vocabulary acquisition during the four weeks.
with \( p < .01 \). This finding complies with what has been reported by researchers, that children have more extensive opportunities to develop their listening proficiency if their instructors select appropriate multimedia materials and assistance (Van Scoter, Ellis & Railsback, 2001; Wright & Shade, 1994). In this study, the use of mobile devices with videos allows the elementary school students to have more extensive opportunities to practice in order to promote their listening comprehension and vocabulary acquisition.

In addition, this study also examined the learning progress of students' listening comprehension and vocabulary acquisition within the four weeks. Figure 4 shows the average test scores of the three groups over a period of a month. It is found that all of the three groups made progress each week; however, the two experimental groups had significantly better learning effectiveness than the control group, in particular, since the third week. In Figure 4, it can also be seen that the two experimental groups revealed similar learning performance during the month.

**Figure 4.** The improvement progress of listening comprehension (left) and vocabulary acquisition (right)

### Difference in learning achievement of diverse learning preferences

This study further investigated the effect of different caption modes on the English listening comprehension and vocabulary acquisition of the students with different learner preferences. Table 3 displays the ANCOVA analysis of vocabulary acquisition which shows a significant difference for the visual learning preference students among the different groups. The students who prefer a visual style of learning in the target-word group performed as well as the students who prefer a visual style of learning in the full-caption group for vocabulary acquisition; moreover, the students who prefer a visual style of learning in both the target-word group and the full-caption group outperformed the students who prefer a visual style of learning in the non-caption group. Accordingly, for the students who prefer a visual style of learning, it is suggested to provide both English and Chinese target words to them, in particular, for those low-achievement students.

**Table 3.** ANCOVA analysis of vocabulary acquisition of the students preferring visual/verbal styles of learning

<table>
<thead>
<tr>
<th>Learning style</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted Mean</th>
<th>F</th>
<th>Pairwise Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal (L)</td>
<td>Control group(L1)</td>
<td>7</td>
<td>51.43</td>
<td>27.95</td>
<td>51.17</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental group one(L2)</td>
<td>9</td>
<td>73.33</td>
<td>33.17</td>
<td>69.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental group two(L3)</td>
<td>9</td>
<td>62.22</td>
<td>38.01</td>
<td>68.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual (V)</td>
<td>Control group(V1)</td>
<td>18</td>
<td>47.37</td>
<td>24.23</td>
<td>51.55</td>
<td>3.23</td>
<td><em>(V1) &lt; (V3)</em></td>
</tr>
<tr>
<td></td>
<td>Experimental group one(V2)</td>
<td>19</td>
<td>70.53</td>
<td>23.45</td>
<td>68.71</td>
<td></td>
<td><em>(V1) &lt; (V2)</em></td>
</tr>
<tr>
<td></td>
<td>Experimental group two(V3)</td>
<td>17</td>
<td>74.12</td>
<td>30.63</td>
<td>70.30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05

Consequently, based on the findings of this study, it is recommended that learners who prefer a visual style of learning use the assistant mode applied in experimental group two for the purpose of promoting their learning comprehension, because they tend to pay attention to the image or video input rather than relying on listening to the
verbal input. In contrast, the students in the target-word group had similar performance in the listening comprehension to the students in the full-caption group. As a result, there is no need to provide full captions, as the use of target words only can provide students with more opportunities to practice listening to various pronunciation features, such as reduced forms, assimilation, elision, and re-syllabification.

Perceptions of the target word subtitles

Nowadays, there is a high level of acceptance of the use of information technology such as mobile devices as learning assistance. This study especially paid attention to the acceptance of experimental groups one and two. Therefore, Table 4 shows the descriptive statistics results of TAM for experimental groups one and two. On the six-point Likert-scale, where 1 represents “strongly disagree” and 6 represents “strongly agree,” the descriptive statistics of the technology acceptance model in experimental group two all attain scores higher than 5 on average. This means that the students agreed with and had positive perceptions of satisfaction, usefulness, and ease-of-use of the mobile learning system embedded with target-word Chinese subtitles and target-word English captions, as well as expressing high acceptance of the approach. There is no remarkable difference between the two experimental groups on the three scales of satisfaction, usefulness, and ease of use, as shown in Table 4. In addition, the students highly accepted the assistance of the target-word Chinese subtitles. The following discussion also presents the opinions of learners provided in the interviews for comparison with the results of the technology acceptance model investigation.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Experimental group one</th>
<th>Experimental group two</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>28</td>
<td>4.83</td>
</tr>
<tr>
<td>Usefulness</td>
<td>28</td>
<td>4.65</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>28</td>
<td>4.58</td>
</tr>
</tbody>
</table>

Discussion and conclusions

This study found that the target-word strategies have better effects on vocabulary acquisition than on listening comprehension for low-achievement elementary school students. As students in non-English speaking countries generally lack English vocabulary, they especially need the assistance of the target words when watching videos for vocabulary acquisition. On the other hand, the results concerning listening comprehension are different from those of previous studies carried out in universities. One possible reason is that undergraduates have learned the frequently-used 2,200 English words, while elementary school students have only learned a few of these words. Consequently, the undergraduates were able to benefit from the practice, whereas the younger students could not (Hsu & Chang, 2010). In conclusion, this study suggests that the partial hidden caption mechanism can be used in an adaptive way that presents the selected vocabulary with different degrees of difficulty based on the learning level of the students.

From the interview results, there are several interesting findings. For example, the students in experimental group one (the full English caption group) indicated that it was not necessary to provide them with full English captions; moreover, they stated that showing full captions interfered with their listening to the learning materials. They believed that providing only target words was sufficient to assist them in improving their listening comprehension, which conforms to the results of the investigation of perceptions of using the system for learning English listening. The results showing the importance of target words and no need for full captions are similar to the research results of Guillory (1999) and Taylor (2005). In addition, some students said that they would like to learn from watching videos and playing computer games related to the topics in their textbooks.

Moreover, most of the students pointed out that the use of mobile devices was another important factor that motivated them to learn. They indicated that such a mobile technology-enhanced learning approach allowed them to practice anywhere and anytime, and hence learning the English listening content seemed to be easier and more interesting. The teacher who participated in this experiment further indicated that "In the beginning, I wondered if the use of mobile devices could benefit the students. I worried that the screen of the PDA might be too small, and the
students might spend lots of time playing games. Surprisingly, the students made significant progress each week, which has changed my perception in terms of using mobile technologies to learn." Such findings are consistent with the experiences reported by Wong et al. (2010).

In the future, we plan to conduct more experiments from several perspectives. The first is to use longer videos, since the present study only applied short videos of 20 seconds. This approach is suitable for elementary school students with low learning achievement, but may not be appropriate for advanced learners. Secondly, the study suggests that future researchers can determine which target words to show, and hide the other words in the captions of different video lengths for learners of different ages. Thirdly, although the results of this study seem to be desirable, it might be difficult to claim that the findings are significant since the test period was so short. It is possible that part of the results may be due to the novelty of viewing the videos with handheld devices. Therefore, it is worth conducting extended studies with a longer test period in the future. It is expected that the extended learning time may make contributions to listening proficiency, especially as there is a trend that mobile devices will become a common learning device in the near future (Norris, Hossain, & Soloway, 2011; Peng et al., 2009). Fourthly, as there are other handheld devices besides PDAs (e.g., e-books, notebooks, smartphones), it would be interesting to compare the learning performance of PDA learners with that of students who learn with other handheld devices. Fifthly, because English teachers may not have the necessary skills in information and computer technologies, we plan to establish a caption processing tool freeware. When English teachers decide to employ this approach in their classes, they can use the freeware themselves to process the captions of their videos, or to select what words they want to show or hide in the captions, such as the target 1,200 words at the elementary and junior high school level, or the Dolch Sight words or keywords. Developing caption-filtering freeware contributes to making captions or subtitles reusable, diversiform and adaptive so that the application of videos in learning via mobile devices can be adapted to different users.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 99-2511-S-011-011-MY3 and NSC 100-2631-S-011-003.

References


Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271-289.


Luo, C. P. (2008). An action research plan for developing and implementing the students’ listening comprehension skills. *English Language Teaching, 1*(1), 25-28


