Educational Technology & Society
An International Journal

Aims and Scope

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

- Educators aim to use technology to enhance individual learning as well as to achieve widespread education and expect the technology to blend with their individual approach to instruction. However, most educators are not fully aware of the benefits that may be obtained by proactively harnessing the available technologies and how they might be able to influence further developments through systematic feedback and suggestions.
- Educational system developers and artificial intelligence (AI) researchers are sometimes unaware of the needs and requirements of typical teachers, with a possible exception of those in the computer science domain. In transferring the notion of a 'user' from the human-computer interaction studies and assigning it to the 'student', the educator's role as the 'implementer/ manager/ user' of the technology has been forgotten.
- The aim of the journal is to help them better understand each other's role in the overall process of education and how they may support each other. The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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


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

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

Editorial Assistant
Barbara Adamski, Athabasca University, Canada.

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

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

Assistant Editors
Sheng-Wen Hsieh, Far East University, Taiwan; Dorota Mularczyk, Independent Researcher & Web Designer; Ali Fawaz Shareef, Maldives College of Higher Education, Maldives; Jarkko Suonen, University of Joensuu, Finland.

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

ISSN 1346-4522 (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@ieee.org.

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

Abstracting and Indexing

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

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

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

Submission procedure
Authors, submitting articles for a particular special issue, should send their submissions directly to the appropriate Guest Editor. Guest Editors will advise the authors regarding submission procedure for the final version. All submissions should be in electronic form. The editors will acknowledge the receipt of submission as soon as possible.

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

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

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

### Special issue articles

- **Guest Editorial – Knowledge Visualization for Learning and Knowledge Management**  
  Minhong Wang and Michael J. Jacobson  
  Pages 1-3

- **Visualizing Topic Flow in Students' Essays**  
  Stephen T. O'Rourke, Rafael A. Calvo and Danielle S. McNamara  
  Pages 4-15

- **Concept Maps as Cognitive Visualizations of Writing Assignments**  
  Jorge Villalon and Rafael A. Calvo  
  Pages 16-27

- **Knowledge Visualization for Self-Regulated Learning**  
  Minhong Wang, Jun Peng, Bo Cheng, Hance Zhou and Jie Liu  
  Pages 28-42

- **Using Computer-Assisted Multiple Representations in Learning Geometry Proofs**  
  Wing-Kwong Wong, Sheng-Kai Yin, Hsi-Hsun Yang and Ying-Hao Cheng  
  Pages 43-54

- **Learning as 'Knowing': Towards Retaining and Visualizing Use in Virtual Settings**  
  Demosthenes Akoumianakis  
  Pages 55-68

### Full length articles

- **Screen-capturing System with Two-layer Display for PowerPoint Presentation to Enhance Classroom Education**  
  Yen-Shou Lai, Hung-Hsu Tsai and Pao-Ta Yu  
  Pages 69-81

- **Metacognitive Process in Online Text Construction**  
  Hui-Chin Yeh and Yu-Fen Yang  
  Pages 82-101

- **A Systematic Approach for Learner Group Composition Utilizing U-Learning Portfolio**  
  Yueh-Min Huang and Ting-Ting Wu  
  Pages 102-117

- **Using Online EFL Interaction to Increase Confidence, Motivation, and Ability**  
  Wen-chi Vivian Wu, Ling Ling Yen and Michael Marek  
  Pages 118-129

- **Investigation of Organizational Interaction and Support in an NGO through Computer-Mediated Discussions**  
  Yao-Jen Chang and Yao-Sheng Chang  
  Pages 130-140

- **A Simulation Model that Decreases Faculty Concerns about Adopting Web-Based Instruction**  
  Hae-Deok Song, Wei-Tsong Wang and Chao-Yueh Liu  
  Pages 141-151

- **Identification of Dysfunctional Cooperative Learning Teams Using Taguchi Quality Indexes**  
  Chin-Min Hsiung  
  Pages 152-162

- **Web-based Spatial Training Using Handheld Touch Screen Devices**  
  Norena Martin-Dorta, Jose Luis Saorin and Manuel Contero  
  Pages 163-177

- **Using Reflective Peer Assessment to Promote Students' Conceptual Understanding through Asynchronous Discussions**  
  Huann-shyang Lin, Zaway-R Hong, Hsin-Hui Wang and Sung-Tao Lee  
  Pages 178-189

- **Supporting Mobile Collaborative Activities through Scaffolded Flexible Grouping**  
  Ivica Boticki, Chee-Kit Looi and Lung-Hsiang Wong  
  Pages 190-202

- **QTIMaps: A Model to Enable Web Maps in Assessment**  
  Toni Navarrete, Patricia Santos, Davinia Hernández-Leo and Josep Blat  
  Pages 203-217

- **Evaluation of the Learning Path Specification**  
  José Janssen, Adriana J. Berlanga and Rob Koper  
  Pages 218-230

- **AUPress: A Comparison of an Open Access University Press with Traditional Presses**  
  Rory McGreal and Nian-Shing Chen  
  Pages 231-239
Building Lectures and Building Bridges with Socio-economically Disadvantaged Students

Peter Phillips and Birgit Loch

Generic Educational Knowledge Representation for Adaptive and Cognitive Systems

Arturo Caravantes and Ramón Galán

240-251

252-266
Guest Editorial - Knowledge Visualization for Learning and Knowledge Management

Minhong Wang¹ and Michael J. Jacobson²

¹Faculty of Education, The University of Hong Kong, Hong Kong // ²Faculty of Education and Social Work, The University of Sydney, Australia // magwang@hku.hk // michael.jacobson@sydney.edu.au

Knowledge visualization is an emerging area, in which visual imagery is used to construct and convey complex insights for improving understanding and communication. Knowledge visualization has been widely used in various learning environments to augment cognition, facilitate thinking and scientific inquiries, and assist meta-cognition. By representing cognitive structures or mental models using various techniques, knowledge visualization has the potential to facilitate the construction of understandings at a deeper level as well as with multiple perspectives such as interpretation and abstraction. Interpretation employs a familiar representation to support or revise misconceptions of the understanding of less familiar or more abstract knowledge. Examples include conveying an abstract idea by relating it to a concrete phenomenon, relating new insights to already understood concepts, and simulating complex objects or processes. Abstraction, on the other hand, refers to “throwing details” by synthesis and generalization. Abstraction exposes the underlying structure of a concept, or abstraction can involve the creation and organization of high-level concepts, actions, and procedures in the form of concept maps (Novak, 1998), knowledge structures, conceptual frameworks, and so on. Knowledge conveyed by abstraction is usually schematic and analytic, with a format that is typically highly structured and systemic due to the reduction of conceptual complexity and the articulation of relationships (Ainsworth, 2006). In addition to deep understanding and high level thinking, knowledge visualization can support meta-cognition by helping learners develop an awareness of their mental representations, which may help foster conceptual change and knowledge transfer (Jacobson, 2004).

In the current global knowledge economy, the importance of learning is seen not just in traditional educational contexts but also in informal and practical work settings as well as in self-directed and lifelong learning. Accordingly, knowledge management (Nonaka & Takeuchi, 1995) is increasingly being recognized by individuals and organizations as an important competitive strategy. In addition to learning or consuming knowledge, people are increasingly becoming engaged in knowledge management processes, such as contributions of new knowledge from practice, applications and verifications of existing knowledge, and development of intellectual assets for improving individual, group, and organizational performance. In these contexts, knowledge visualization offers great potential for the creation of new knowledge from individual or group activities using heuristic sketches or rich graphic metaphors. Innovation or knowledge creation can be fostered by making preliminary ideas or unstable knowledge explicit and debatable. Moreover, visualization improves knowledge application or problem solving by providing rich representations to express “know-how” knowledge exercised in the accomplishment of tasks (Wang, Ran, Liao & Yang, 2010), which is usually tacit and associated with hands-on experience and practice at solving problems. Visualizations may externalize or reify tacit know-how knowledge into explicit formats such as competency models, workflow charts, business process diagrams, decision-making models, reasoning patterns, and argument processes.

Another area closely related to knowledge visualization is information visualization (Tufte, 1990), which is the visual representation of datasets. Information processing and knowledge construction are closely intertwined in the learning process as learners need to access information to acquire knowledge and formulate understandings. While knowledge visualization is primarily intended to facilitate subject understanding and interpretation, information visualization supports the perception of patterns and structural relations in data via data manipulation, data analysis, data representation, and data mining techniques. In addition, information visualization is becoming more socially constructed, especially in web-based environments, which in turn has generated work on social visualizations for finding trends and patterns of people and their behavior in social communities (Gilbert & Karahalios, 2009). Representations like network diagrams, tag clouds, tag clusters, and scatterplots for social visualizations help foster an awareness and accountability of a community’s behavior and structure, thus serving as a medium for reflection, motivation, and community contributions.

Various methods or techniques from paper-pencil to computer-based tools have been used to create images, diagrams, or animations to communicate both abstract and concrete ideas. In particular, computers are used as electronic tools for reflecting human cognitive processes through visual representations on the screen. Computer-
based visual representations may amplify, extend, or enhance human cognitive functions, thus engaging users as they represent, manipulate, and reflect on what they know (Jonassen, 2000). Moreover, techniques for dynamic and interactive visualization are being explored and utilized (Lowe, 2004). While there is no doubt that technologies enable and promote knowledge visualization in various ways, the educational benefits of knowledge visualization do not occur automatically. To foster effective learning and knowledge construction, pedagogical functions should be integrated with visual representations, and cognitive tasks must be undertaken and scaffolded in many learning processes aimed at understanding complex phenomena (Spector, 2006).

Knowledge visualization, viewed broadly to include both information and social visualization, has potential advantages for encouraging deeper understanding, hypothesis building, reasoning, and problem solving, and these advantages have been well recognized in learning and knowledge management research and practice. However, as an emerging area, there have been relatively few studies that address the multifaceted nature of knowledge visualization and its role in representing and organizing knowledge as well as in supporting learning. The articles in this special issue address some of the broad issues discussed above.

In the first paper “Visualizing Topic Flow in Students’ Essays” by Stephen T. O’Rourke, Rafael A. Calvo, and Danielle S. McNamara presents a document visualization technique and a measure of document quality based on the semantic distance between the parts of a document. The visualization in this study focuses on the structure and the flow of topics of a document. The results provide evidence that the degree of topic flow between consecutive sentences and paragraphs is consistent with the essay quality and helps assessing and improving the quality of essay writing.

Concept mapping as an important tool for knowledge visualization has been widely used to help learners construct, communicate, and assess their understanding. Jorge Villalon and Rafael A. Calvo proposed to extend the use of concept maps from reading to writing by proposing a concept miner tool to automatically generate concept maps embedded in student writings. Their paper “Concept Maps as Cognitive Visualizations of Writing Assignments” presents the design and implementation of the tool and its integration with a writing support environment that provides automatic feedback to students in their writing activities.

While concept mapping is effective in fostering in-depth thinking and understanding, self-constructed maps pose high cognitive demands on learners’ ability to integrate multiple forms of thinking for knowledge analysis and externalization. The paper “Knowledge Visualization for Self-Regulated Learning” by Minhong Wang, Jun Peng, Bo Cheng, and etc. proposes the use of knowledge maps created by experts to guide and scaffold the self-directed learning process of novices who suffer from “lost-in-hyperspace.” The focus of their study is on visualizing the domain knowledge structure and integrating it with the curriculum, learning resources, intellectual process, learning assessment, and social learning process.

To support learning in different ways, knowledge can be visualized with multiple representations. Wing-Kwong, Wong, Sheng-Kai Yin, Hsi-Hsun Yang, and etc. explored the use of computer-assisted multiple representations in learning geometry proofs. Their paper “Using Computer-Assisted Multiple Representations in Learning Geometry Proofs” investigates the design of multiple representations to support the learning of theorem proving. Their visual representations cover problem description, static and dynamic geometry figures, formal proof, and proof tree, which are designed for different pedagogical functions and integrated with a computer-assisted learning environment.

In the last paper of the special issue, Demosthenes Akoumianakis looked into the visualization of collaborative practices with electronic traces in virtual settings. His paper “Learning as ‘Knowing’: Towards Retaining and Visualizing Use in Virtual Settings” investigates the role of and challenges for visualization in mediating collaborative activities and leading to recurrent co-engagement in practice. Two case studies were presented where visualizations are employed as sharable and accountable practice vocabularies capable of capturing collective intelligence, thereby unfolding a variety of informal learning patterns.

We conclude by noting that the papers in this special issue are intended to be representative of ongoing research in the area of knowledge visualization for learning and knowledge management. The international scope of this research is distinctive, and the breadth of applications for knowledge visualization discussed in these papers is important. Overall, we hope this special issue and the individual research projects that are featured will foster further interest in what we believe will become an area of increasing importance as new visualization techniques are
developed and their efficacy explored to support individual and collective learning and understanding in schools and organizations.

References


Visualizing Topic Flow in Students’ Essays

Stephen T. O’Rourke, Rafael A. Calvo and Danielle S. McNamara

University of Sydney, Australia // 1University of Memphis, USA // stephen.orourke@sydney.edu.au // rafael.calvo@sydney.edu.au // dsmcnamara1@gmail.com

ABSTRACT
Visualizing how the parts of a document relate to each other and producing automatically generated quality measures that people can understand are means that writers can use to improve the quality of their compositions. This paper presents a novel document visualization technique and a measure of quality based on the average semantic distance between parts of a document. We show how the visualization helps tutors mark essays more efficiently and reliably, and how a distance index calculated for the visualizations correlates with grades. The technique is further evaluated using three dimensionality reduction techniques. The results provide evidence that the degree of topic flow between consecutive sentences and paragraphs is related to essay quality.

Keywords
Document visualizations, writing, argument flow

Introduction

Writing is an important learning activity common at all educational levels and disciplines. But incorporating writing activities into the curricula faces many challenges, including the cost of providing meaningful and timely feedback as the one provided in assessment. Technically researchers are tackling these challenges by producing automated feedback (including grades) directly targeted to the students or support for human assessors who then write the feedback. In this line of research studies have shown tools and techniques for automated feedback in academic writing (Beals, 1998; Graesser & McNamara, in press; Thiesmeyer & Thiesmeyer, 1990; Wade-Stein & Kintsch, 2004; Wiemer-Hastings & Graesser, 2000). Feedback can be genre specific, as for example, in argumentative writing. Many studies in this area have focused on argument visualization (Kirschner, Shum, & Carr, 2003) where the students are visually presented with the way in which claims are interrelated, showing evidential structures. Other forms of feedback focus on quality measures that apply to multiple genres and disciplines. For example, automatically generated visualizations can be used as support material that students can use to reflect on a set of trigger questions designed around issues with which students normally have difficulty (Calvo & Ellis, 2010). Thanks to new cloud computing technologies these computationally intensive forms of support can be provided in real-time, at any stage of the writing process and to large numbers of students (Calvo, O’Rourke, Jones, Yacef, & Reimann, 2011).

Many different features of a document can be quantified and therefore represented visually as a form of feedback or support to handling writing activities. The challenge is to find visual representations of features that are both meaningful to actual writing tasks (e.g. putting together evidence into an argument) rather than mathematical artifacts, and that do actually provide useful information that correlates with the quality of writing. Several linguistic features of quality writing have been identified with evidence that they can predict high and low proficiency essays. For example, McNamara (2010) provided evidence for syntactic complexity (measured by the number of words before the main verb), lexical diversity, and word frequency (as measured by Celex, logarithm of all words). While other measures, such as cohesion indices, have received much attention in the literature, they generally offer more of a guide to a text’s connectedness, which does not necessarily correlate well to that of experts (Crossley & McNamara, 2010). The way an argument is structured and the flow in a composition are important quality features. We evaluate here different techniques for producing topic flow visualizations. These were proposed by O’Rourke and Calvo (2009b) as a way of helping students reflect about their writing, particularly issues related to the flow in the composition. The visualization techniques include several steps, all of which must be taken into account if the actual visual representation (the final outcome) is to be semantically valid and useful.

In most Natural Language Processing techniques, each text segment is converted into a high dimensional representation using the Vector Space Model. This high dimensional space is then reduced by using one of several mathematical techniques that preserve as much of the original information as possible. The first challenge is to find the optimum dimensionality reduction technique that produces meaningful visualization. The optimum value is the measure of how it relates to a writing quality attribute (e.g. flow). This reduced space can then be used to produce a
2-dimensional space that can be made into a visual representation. In our topic flow visualizations, textual units (e.g. paragraphs) are represented in a two dimensional space, with their distances being proportional to the distance between topics (in the reduced dimensionality space). This visualization can provide the means to see how the parts of a document interrelate and follow each other. Once the vector data is in lower dimensional representation and can be represented visually we need to understand how the visualization (produced using optimum dimensionality reduction techniques) will improve an aspect of a learning activity such as reading and assessment. This is the second important question addressed in this paper.

The new techniques for document visualization and for quantifying topic flow are described in the next section. First the mathematical framework to represent text segments as vectors (i.e. the vector space model), and then the combination of the three dimensionality reduction techniques and Multidimensional Scaling, used to bring these representations to a 2-dimensional visualization. Since the algorithms need to be validated as producing data that represents the actual semantics of the documents we show how they can be used to quantify topic flow, in a collection of real documents assessed by experts. We then explore the value of the actual visual representations by measuring the impact they have on tutors assessing a collection of essays. This evaluation is made on a new corpora of student essays and assesses to what extent semantic features can be captured using these techniques.

**Visualizing Topic Flow**

**The Mathematical Framework**

Documents can be represented in high dimensional spaces (e.g. all its possible topics or terms) but their visualization can only be done in 2 or 3 dimensions. It is therefore essential to choose the dimensions that are meaningful representations of quality features found in the text. Textual data mining approaches have been applied to analyze the semantic structure of texts (Graesser, McNamara, Louwerse, & Cai, 2004) and to build automatic feedback tools (Villalon, Kearney, Calvo, & Reimann, 2008; Wiemer-Hastings & Graesser, 2000). Such approaches typically involve performing computations on a text’s parts to uncover latent information that is not directly visible in the surface structure of text. These approaches have been applied to solve a number of problems in educational settings, including automatic summarization, automatic assessment, automatic tutoring and plagiarism detection (Dessus, 2009). These applications often use measures of semantic similarity to compare text units. The most common approach involves the creation of a term-by-document matrix; derived from frequency vectors of distinct terms in each document to create a topic model, which describes the mixture of different topics in the document corpus.

The most representative dimensionality reduction techniques (a.k.a. topic models) are Latent Semantic Analysis (LSA) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), Non-negative Matrix Factorization (NMF) (Lee & Seung, 1999), and Probabilistic LSA (PLSA) (Hofmann, 2001). These dimensionality reduction techniques produce topic models, where individual terms in the term-by-document matrix are weighted according to their significance in the topic (McNamara, 2010). The aim of dimensionality reduction is to eliminate unimportant details in the data and to allow the latent underlying semantic structure of the topics to become more evident. These topic modeling methods do not take into consideration many important components of a text, such as word order, syntax, morphology, or other features typically associated with text semantics, such as linking words and anaphora. Nevertheless, the topic models built using these methods from many source documents have long been shown to be reliable in the domain of information retrieval and highly similar to topic models extracted by humans (c.f.p. Landauer, McNamara, Dennis, & Kintsch, 2007). Similarly, automatic essay scoring systems have also been shown to have agreement comparable to that of human assessors with the use of pre-scored essays (Wang & Brown, 2007) and course materials (Kakkonen & Sutinen, 2004) to build the mathematical models.

While the above-mentioned topic modeling techniques have been successfully used on a large scale, they can be somewhat problematic in the case of providing feedback on a single essay. The corpus used to create a topic model is essentially a set of baseline texts, which are used to define the topic semantics against which a document will be compared. Thus, the distances between the term vectors in a semantic space of a topic model is entirely dependent on the corpus upon which it is built. A document is compared to how it relates to the tendency words to co-occur in the source documents, which could bias its meaning.
Another approach is to create the topic model from only the information contained in the document itself. This would maximize the variance of the model, while focusing on the semantics of the actual document itself, rather than relating it to something external. The idea of constructing a topic model from a single document was first proposed by Gong and Liu (2001) for the purposes of automatic text summarization. The single document semantic space technique has also been used by other researchers for automatic summarization (Steinberger, Poesio, Kabadjov, & Jeek, 2007) and discovering labels for clustered results in information retrieval (Osinski, 2006). More recently, Villalon and Calvo (2009) showed that the rank order of document distances in a topic model created from a single document was similar to one created from many source documents.

The visualization technique introduced here involves performing the following steps. First, a term-by-sentence matrix is prepared, after stop-words and low frequency words are removed, and stemming is applied. Second, a topic model is created using the NMF dimensionality reduction technique. Third, the topic model is projected to a 2-dimensional space using Multidimensional Scaling, and finally, a visualization of the document’s paragraphs is produced. For brevity, the details of the three techniques used are left to their original references.

The elements of the initial term-by-paragraph matrix can be weighted using a number of schemes (e.g. Chi-squared, Log-entropy, TF-IDF) (Baeza-Yates & Neto, 1999). Log-Entropy, used in this paper, weighs a term by the log of its frequency $tf_{ij}$ in a paragraph offset by the inverse of the entropy of its frequency across all $n$ paragraphs in a document. The formula for calculating the Log-entropy weight of a term entry is defined in equation (1). Log-Entropy provides a useful weighting scheme for our purposes because it assigns higher weights to terms that appear fewer times in a smaller number of paragraphs. Thus, emphasizing the importance of infrequent terms in the paragraphs while also eliminating the ‘noise’ of frequent terms.

$$x_{ij} = \frac{\log (1 + tf_{ij})}{-\sum_{k=1}^{n} \frac{tf_{ik}}{\sum_{l=1}^{n} tf_{il}}} \log \left( \frac{tf_{ik}}{\sum_{l=1}^{n} tf_{il}} \right)$$

The NMF dimensionality reduction technique generates its topic model by decomposing $X$ into the product of two $k$-rank non-negative matrices, $W$ and $H$, so that $X$ is approximately equal to $X \approx WH$. In our case, $k$ is considered to be the number of latent topics in a document. This makes the choice of $k$ entirely document dependent. Given that $k$ represents the number of latent topics in a document, $W$ becomes a term-by-topic matrix, indicating the weighting of each term in a topic, and $H$ becomes a topic-by-document matrix, indicating the weight of each topic in a document. The product $WH$ is called a nonnegative matrix factorization of $X$ which can be approximated by minimizing the squared error of the Frobenius norm (Meyer, 2000) of $X - WH$. Finding this solution defines the NMF problem which can be mathematically expressed as:

$$F(W, H) = \|X - WH\|^2_F$$

The distance between any two paragraphs (not only the consecutive ones) can be calculated in the reduced topic representation using standard measures (cosine, Euclidean, etc.). Multidimensional Scaling uses this paragraph-paragraph triangular distance table to produce a 2-dimensional representation (Borg & Groenen, 2005). The Multidimensional Scaling transformation is performed using a procedure called iterative majorization (de Leeuw, 1977). The iterative majorization algorithm undertakes a least-squares approach to Multidimensional Scaling by attempting to minimize a loss function called Stress. Stress (Equation 3) can be expressed as the normalized sum of the squared errors between the vector dissimilarities $d_{ij}$ and their approximated distances $\hat{d}_{ij}$ in the low dimensional space. The final result is a least-squares representation of the paragraphs described in the distance matrix, with the directions of the axes being arbitrary.
Visualizing Flow

A well-structured essay should have a clear and logical flow of ideas represented through its flow in paragraphs. The 2-dimensional visual representation can be used by students to reflect on the correctness of each argument (i.e., how each argument point follows each other). The visualization may also help detect ‘breaks’ in the flow, which is when two consecutive paragraphs talk about very different topics. The visualization presents the document in a different way, as it might appear for an external reader. In a paragraph ‘map’ such as the one in Figure 1, the essay’s paragraphs are plotted on a circular grid with the diameter of the grid equal to the maximum possible distance between any two paragraphs (i.e., no topic overlap). The paragraphs are represented using a node-link diagram with text labels and arrows used to indicate the paragraph sequence.

For example, the clear sequence of topics in the five-paragraph essay paradigm (Davis & Liss, 2006), can be visualized in our map. In this particular genre, the content of the ‘introduction’ and ‘conclusion’ paragraphs is expected to be similar, so these paragraphs should appear close in a map. The ‘body’ paragraphs address different subtopics and should ideally be linked through transitions so they should be sequentially positioned in the map. The map of a well-structured ideal five-paragraph essay would have a circular layout of sequential paragraphs, indicating a natural change in topic over the essay, with the introduction and conclusion starting and finishing on similar points. In contrast, we would expect a poorly structured essay to have many rough shifts in topic, with paragraphs positioned somewhat randomly around the map.

The visualization should also make evident the difference between a lower and a higher quality essay. Figure 1 illustrates the paragraph maps of two short essays. The essay on the left was given a low grade while the essay on the right was given a high grade. The topic flow of the high-grade essays clearly resembles that of the prototypical five-paragraph essay described above, while topic flow of the low-grade essay appears disorganized. This qualitative evaluation implies that how much does statistically/geometrically the map/path deviate from a circle indicates the quality of its structure.

\[
\sigma = \sum_{i<j} \frac{(d_{ij} - \bar{d}_{ij})^2}{\bar{d}_{ij}^2}
\]
Quantifying Topic Flow

The topic flow analysis approach uses text mining techniques to model the topic mixture of a document’s sentence and paragraphs, followed by document similarity comparisons to interpret the text’s structure and flow. The automated approach involves performing the following steps:

- First, a term-by-sentence matrix is prepared, after stop-words and low frequency words are removed, and stemming is applied.
- Second, a topic model is created using dimensionality reduction techniques (NMF, SVD, PLSA).
- Finally, similarity comparisons are performed to identify features in the topic model of the document.

The approach is based on techniques that are well established and proven in the literature as well as the theoretical suitability of these techniques for the application. It is designed to be applied to most problems without parameter tuning, or substantial work on stopword lists. In the pre-processing step, after a given English document is decomposed into individual paragraphs, a list of 425 words from the Brown stopword list are removed, and stemming is performed using the Porter stemming algorithm. A term-frequency vector for each paragraph in the document is then constructed from the terms’ stems.

The core of the approach is in the dimensionality reduction, which is used to model the topic flow of a document’s text passages based on their associated topic mixtures. A separate topic model is built for each document, using a document’s content to generate the term-by-sentence matrix. In the case of SVD and NMF, the matrix elements of the initial term-by-sentence matrix can be weighted using a number of schemes (Baeza-Yates & Neto, 1999).

Once the topic model is built, each topic is represented as a vector of its distribution of terms and each sentence is represented as a vector of its distribution of terms over these topics: thus producing the topic model from which an analysis of a document’s semantic structure and flow can be performed. However, actually quantifying the topic flow in an essay is a difficult task and a methodology for doing so is missing from the literature. While a break in topic flow can sometimes be a good thing, on average it is reasonable to expect that a well written essay’s topic flow should be better than that which would be expected from random chance.

In the context of the research presented in this paper, topic flow is quantified as the average amount of semantic overlap between successive sentences or paragraphs in an essay (O’Rourke & Calvo, 2009a). The Distance Index (DI), defined in Equation (4), measures the sum of semantic distances \( \hat{d}_{ij} \) between consecutive pairs of sentences or paragraphs, ‘centered’ and normalized by the average over all the pairs of sentences or paragraphs in a document. These averages are equivalent to distances that would be expected from randomizing the order of the paragraphs. A DI value less than or equal to 0 indicates a random topic flow, while a DI value greater than 0 indicates the presence of topic flow.

\[
DI = 1 - \frac{\sum_{i=1}^{n-1} \hat{d}_{i+1}}{\frac{1}{2} \sum_{n} n \sum_{i<j} \hat{d}_{ij}}
\]

Evaluation 1: Flow and Grades

Although the performance of various dimensionality reduction techniques has been examined by many researchers in the literature, the differing results have exposed how the suitability of each technique seems to be domain-dependent. Our approach of measuring the semantic flow of a single document with the use of a semantic space created solely from the actual document itself is unique, and makes the algorithm independent of any particular domain knowledge. This section evaluates the NMF, PLSA and SVD dimensionality reduction techniques in measuring the semantic flow of a text.
Experiment Dataset

The evaluation was performed using a corpus consisting of 120 essays written for assignments by undergraduate students at Mississippi State University (McNamara, et al., 2010). The corpus includes its associated essay grades from which a quality judgment about the essay semantics is inferred. The essays have been graded from 1 to 6. The essays contain an average of 726.20 (SD=114.37) words, 40.03 (SD=8.29) sentences, and 5.55 (SD=1.32) paragraphs.

Measuring Topic Flow

The aim of this experiment was to validate quantitatively whether dimensionality reduction techniques can be used to analyze the topic flow of an essay and measure whether the choice of dimensionality reduction techniques used in the approach is in line with the theoretical justifications discussed in the background sections. As topic flow is generally considered a positive feature of an essay, the assumption was made that these essays do have a measurable degree of topic flow, and that high graded essays have a higher amount of topic flow than that of low grade essays. In order to quantify topic flow the Distance Index (DI) measure in Equation (4) was used.

For each essay, a term-by-sentence weight matrix was calculated using the log-entropy term weighting scheme (other schemes had similar results). Since different dimensionality reduction techniques may affect the DI measure, this evaluation compares the results of the Singular Value Decomposition (SVD), Non-negative Matrix Factorization (NMF), and Probabilistic Latent Semantic Analysis (PLSA) dimensionality reduction methods. The number of dimensions (topics) used for the dimensionality reduction algorithms was kept at \( k = 5 \) throughout the experiment. This parameter was chosen based on experimental experience with this dataset; however, similar results were achieved with other \( k \) values. The ideal of number of \( k \) dimensions is entirely document dependant and automatically determining this value is out of the scope of this paper.

The distance between the pairs of sentences and pairs of paragraphs in the reduced semantic space was calculated using the measure of cosine similarity. The DI was used to calculate and compare the difference in topic flow between the sentences and paragraphs for the graded essay subsets produced using the different dimensionality reduction methods.

Two evaluations were performed. In the first, the essay corpus was divided into a low grade and a high grade subset in order to define a quality benchmark on which to critically evaluate the experimental results. The low grade subset consisted of 67 essays graded 1 to 3 and the high grade subset consisted of 53 essays graded 3.2 to 6. In the second evaluation a correlation coefficient between DI and grades was calculated.

Results

The results for the sentence topic flow and the paragraph topic flow in the MSU corpus are summarized in Table 1. A graph illustrating the difference in the average sentence DI and the average paragraph DI produced using the SVD, NMF, and PLSA dimensionality reduction techniques is displayed in Figure 2.

The average sentence DI was found to be relatively larger in the high grade subset than the low grade subset for all the evaluated dimensionality techniques. The difference in the sentence DI between the graded essay subsets was found to be statically significant using NMF and SVD (\( p<0.05 \)), but not PLSA (\( p>0.05 \)).

The average paragraph DI, on the other hand, was less present using either of the dimensionality reduction techniques, with a value of close to 0 and little difference between the graded essay subsets. Although the high grade essay subset had a slightly higher average paragraph DI than the low grade subset, the results were not found to be statistically significant (\( p > 0.05 \)) using any of the dimensionality reduction techniques.

<table>
<thead>
<tr>
<th>Table 1. The mean sentence and paragraph distance indexes (DI) produced using the NMF, PLSA and SVD dimensionality reduction techniques</th>
<th>Low grade</th>
<th>High grade</th>
<th>p-value</th>
<th>Effect size</th>
</tr>
</thead>
</table>
However, the measure of statistical significance does not take into consideration the actual size of the effect that topic flow has on an essay grade. This can be quantified by calculating the ‘effect size’, which, as the name suggests, measures the actual size of the difference between two datasets. At the sentence level, the Cohen’s effect size of the DI between the graded essay subsets was calculated to be slightly larger using NMF (0.39) and SVD (0.39) compared to that of PLSA (0.23). This means that NMF and SVD were able to detect a greater effect of topic flow on a grade compared to that of PLSA. At the paragraph level, NMF had the highest effect size (albeit non-significant) followed by PLSA and SVD, with values of 0.17, 0.16 and 0.12, respectively.

Figure 2. A comparison the average sentence and paragraph distance indexes (DI) produced using the NMF and PLSA and SVD dimensionality reduction techniques for the MSU corpus

With such a strong indication of sentence topic flow in the MSU corpus and the availability of numerical scores, the corpus was further analyzed in more fine-grained detail. Results in Table 2 show that on average the sentence DI of the essays increased with the score, with the results for the NMF algorithm being the most consistent. This result is in support the assumption in this paper of topic flow and essay quality. On average, there was a higher amount of topic flow between sentences than paragraphs.

Table 2. Correlation of the sentence and paragraph distance indexes (DI) with the numerical essay scores

<table>
<thead>
<tr>
<th></th>
<th>NMF</th>
<th>PLSA</th>
<th>SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.23 (0.14)</td>
<td>0.21 (0.13)</td>
<td>0.16 (0.13)</td>
</tr>
<tr>
<td>Pearson’s r</td>
<td>0.27</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>Kendall’s τ₀</td>
<td>0.18</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>Spearman’s ρ</td>
<td>0.25</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>Paragraph</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>-0.01 (0.17)</td>
<td>0.02 (0.14)</td>
<td>0.19</td>
</tr>
<tr>
<td>NMF Avg DI</td>
<td>0.21 (0.13)</td>
<td>0.26 (0.15)</td>
<td>0.02</td>
</tr>
<tr>
<td>PLSA Avg DI</td>
<td>0.19 (0.13)</td>
<td>0.22 (0.14)</td>
<td>0.12</td>
</tr>
<tr>
<td>SVD Avg DI</td>
<td>0.13 (0.11)</td>
<td>0.19 (0.15)</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Evaluation 2: Supporting Assessment

Achieving good consistency in the grading of written assignments is difficult and adds to effort to the already time-consuming task. Such consistency is even more difficult when the tutors assessing the assignment are not trained in teaching writing or are non-native speakers (a common scenario). The aim of this evaluation was to study the use of the topic flow maps as a visual aid in the characterization of the texts’ semantic structure and flow, supporting the instructors who assess the quality of the composition.

Methodology

The assessment of essays is a subjective task and there is evidence that the visualization of qualitative data can be used to enhance the speed and accuracy of a subjective task (North, 2006). The study evaluated the time tutors take to complete the marking of a collection of essays and the inter-rater agreement that the tutors had with two expert raters. The two tutors independently marked assignments with and without the visual aid of the topic flow maps. Following North’s work (North, 2006), it is hypothesized that the structure and development of an essay can be subjectively assessed faster, more accurately, and more consistently with the visual aid of a topic flow map.

<table>
<thead>
<tr>
<th>Paragraph</th>
<th>Mean (SD)</th>
<th>Pearson’s r</th>
<th>Kendall’s τb</th>
<th>Spearman’s ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00 (0.16)</td>
<td>-0.02 (0.19)</td>
<td>0.02 (0.18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.05</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.12</td>
<td>0.05</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.06</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The main interface of the essay assessment tool. An essay (Left) is displayed side-by-side with its corresponding map (Right)

The two tutors marked the assignments using the interface shown in Figure 3 (or a version without the visualization). Each of the essay paragraphs was numbered, so they could easily be identified in the map. The rubrics for marking...
are shown at the top of the interface. The rubric focused on questions like: ‘Do you start with an introduction? Is there a clear flow of ideas in your essay? Is there a conclusion that follows from the rest of the essay? Do you make paragraph breaks in places that reflect the structure of your essay?’ Below each question is a list box for the participant to record their score. Each time a participant changes the score in the list box, their assessment time is updated and logged by the tool.

**Essay Subset Preparation**

The evaluation was performed using a corpus of N=43 short essays handwritten in a timed assessment by students at the University of Sydney and then assessed by to expert applied linguistics. The essays were then typed in for this study. Due to the timed nature of the assessment and the fact that essays were originally handwritten, they are often quite erroneous. The corpus includes its associated numerical essay marks, as assessed according to the MASUS procedure (Bonanno, Jones, & University of Sydney. Learning Centre., 2007). Compared to the MSU corpus, the essays in the MASUS corpus are slightly shorter in length, with documents having an average of N=445.85 (SD=120.91) words, N=23.30 (SD=5.04) sentences, N=5.40 (SD=1.74) paragraphs.

Essays with only one or two paragraphs were excluded from the experiment, due to their lack of enough useful semantic information. The 40 essays remaining were divided into two subsets of 20 essays each, with roughly equal distributions of words, paragraphs and scores.

| Table 3. A comparison of the statistical differences between the MASUS essay subset with maps and the MASUS essay subset with no maps |
|---|---|---|
| **Essays** | **Map** | **No map** |
| Total | 20 | 20 |
| **Words** |  |  |
| Mean (SD) | 455.80 (116.17) | 435.85 (127.71) |
| **Paragraphs** |  |  |
| Mean (SD) | 5.60 (1.50) | 5.75 (2.05) |
| **Rating** |  |  |
| Mean (SD) | 3.15 (0.65) | 3.20 (0.59) |

**Results**

The essays were assessed by two tutors with an engineering background, one a native English speaker (Rater 1) and the other a non-native English speaker (Rater 2). The average time taken for the raters to assess the essays as a function of condition is provided in Table 4. In order to eliminate the effect of essay length, the time taken to assess an essay was normalized to a measure of words per second. To eliminate interactions, the tutors did not discuss their work during the task. On average, both raters assessed the essays faster with the visual aid of the maps than without. This is an important result, because the raters are processing more rather than less information when they are provided with the maps, but they processed more words per second in the essays when they were provided with the maps.

| Table 4. A summary of time taken for the two raters to score the MASUS essays, both with and without the visual aid of the topic flow maps |
|---|---|---|
| **Rater 1** | **Map - Time (words/sec)** | **No map - Time (words/sec)** | **p-value** |
| Mean (SD) | 4.60 (0.85) | 3.24 (0.55) | < 0.01 |
| **Rater 2** | **Map - Time (words/sec)** | **No map - Time (words/sec)** | **p-value** |
| Mean (SD) | 4.39 (2.28) | 3.64 (0.66) | 0.08 |

Separate statistical comparisons for each rater revealed that the difference in the time taken to assess the essays with and without the maps for Rater 2 was not statically significant (p = 0.08), whereas it was for Rater 1 (p < 0.01). The
large variability (SD=2.28) in the time taken for Rater 2 to score the essays with the aid of the maps can be explained by an outlier entry in the data. However, excluding this entry from the analysis did not improve the statistical significance of the result.

The most important dimension for evaluating whether the raters gained an insight into the semantics of the essay using the maps is by measuring how the maps affected the accuracy and consistency of the scores. This can be measured by calculating the correlation of the rater scores to that of the expert. Correlation determines the inter-rater agreement between two raters. In order to obtain a comprehensive view of this dimension of the results, both nominal and ordinal correlation measures are used.

Using Pearson’s r, the respective inter-rater agreement with the expert MASUS raters for Rater 1 and Rater 2 was $r = 0.69$ and $r = 0.63$ using the maps compared to $r = -0.14$ and $r = -0.02$ without the maps. In both cases, there was a large different in Pearson’s correlation for the two essay conditions. Therefore, both raters used the rating scale more consistently in the essay condition with the visual aid of the maps.

### Table 5. A summary of the scores given by the two raters and their correlation with the expert MASUS raters

<table>
<thead>
<tr>
<th></th>
<th>Score - Map</th>
<th>Score - No map</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rater 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.25 (0.79)</td>
<td>3.00 (0.92)</td>
</tr>
<tr>
<td>Kappa $\kappa$</td>
<td>0.47</td>
<td>-0.07</td>
</tr>
<tr>
<td>Pearson’s r</td>
<td>0.69</td>
<td>-0.14</td>
</tr>
<tr>
<td><strong>Rater 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>2.35 (1.09)</td>
<td>2.15 (0.83)</td>
</tr>
<tr>
<td>Kappa $\kappa$</td>
<td>0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td>Pearson’s r</td>
<td>0.63</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

### Discussion

This paper aimed to show that the techniques used to compute these semantic spaces, and the flow in a document, can be used to 1) quantitatively predict the quality of a composition 2) produce visualizations that help readers (i.e. tutors) assess qualities of the composition more reliably and faster.

The study evaluated cognitive visualizations of student essays, particularly of the composition’s flow as measured by distances between text units (sentences or paragraphs). We evaluated the correlations between the distances and the essay scores (i.e. quality) as well as the impact of providing the maps to raters on the time and reliability of assessing the quality of essays. This topic flow visualization method involves a process of dimensionality reduction to uncover topics in an essay, followed by multidimensional scaling to map the topic closeness of the essay’s paragraphs to a 2-dimensional representation.

We compared different dimensionality reduction techniques and document similarity comparison for quantifying the semantic features of an essay, and evaluated how these techniques could be used to capture the topic flow of essays. The approach of quantifying topic flow was evaluated using a corpus of 120 essays written by university students. Three dimensionality reduction techniques, NMF, PLSA and SVD were evaluated with respect to the amount of measurable topic flow according to a defined distance index.

The results of this study indicate that NMF, PLSA and SVD capture some effects of semantic flow, which is likely due to greater argument overlap by literal repetition of topics from sentence to sentence. All of the dimensionality reduction techniques indicated that on average they were able to capture the degree to which consecutive sentences discuss similar semantic content. The average distance index between consecutive sentences was found to be greater in the high grade subset than the low grade subset for all the evaluated dimensionality techniques. While the effect size of topic flow between consecutive sentences on an essay grade was calculated to be roughly the same for NMF and SVD, but less for PLSA.

At the paragraph level, however, there was found to be little topic flow between consecutive paragraphs. The effect size of the topic flow on the essays grade was also small. This can be at least partly attributed to the size of the
documents used to build the semantic space. In the experiments of O'Rourke and Calvo (2009a), where much larger documents were used (averaging of 29.42 paragraphs, compared to 5.55 in the present experiment), the paragraph distance index was more closely tied to essay grades. In shorter essays it is likely that each topic will be more confined to individual paragraphs. Similarly it is likely that the topic content in the introduction and conclusions paragraphs will actually be more similar compared to their connecting paragraphs. These features could be observed in many of the topic flow visualizations produced from higher quality essays; where the introduction and conclusion paragraphs were often positioned in close proximity and more central to the other body paragraphs, which were generally positioned more equidistant to each other.

It is clear that the semantic space and thus the inter-document distances differ from one essay to the next, but the challenge remains in providing an automatic means of scoring coherence. The results of this study indicated that there is a measurable amount of topic flow in an essay that relates to the quality of an essay. The results showed that higher quality essays in the corpus had a higher amount of semantic overlap from sentence to sentence, but little from paragraph to paragraph. McNamara et al. (2010) and Crossley and McNamara (2010) have found that cohesion markers are not related to humans’ estimates of essay coherence. Here we have examined the extent to which semantic flow is related to essay quality. There is some evidence that there is a relation, but in this case, the coherence resided primarily at the inter-sentence level, rather than between paragraphs (O'Rourke & Calvo, 2009a).

In the second study, topic flow visualization was evaluated using a corpus of 40 essays written by university students, which had been previously assessed by two expert linguists according to the MASUS procedure. The evaluation demonstrated the use of the visualization for assessing the structure and flow of an essay. On average the experiment participants were shown to assess the essays faster and more accurately and consistently with the aid of topic flow visualization.

Conclusions

Understanding the semantic space of an individual document and determining how its features relate to a grade is in no way simple. Intuitively, the semantic space of a document tells us something about the structure and the flow of topics within the document. How to best quantify this intuition using only the information contained in the document itself remains an open research question. A break in topic flow is not necessarily good or bad, but the semantic space should show that 1) the topic will shift such that it is likely that consecutive parts of a text should be more similar compared to those which are further away 2) there should still be a measurable semantic structure in an essay. Our Distance Index results show that this can be achieved, and they also highlight the impact of different feature selection techniques.

The evaluation also shows that the visualization can be a useful tool for assessment. We found that two raters’ assessment of the essays was more accurate and consistent with the visual aid of the maps. The results are important because the raters assessed the essays more quickly but also more consistently with the expert raters. This has important educational implications because students may be able to better judge their own essays using such a tool and teachers may be able to score essays more quickly and consistently. Clearly more evidence is needed prior to making firm conclusions and further experiments are needed with more raters on various corpora. Nonetheless, these promising results highlight the need for further research on the use of visualization for interpreting essay semantics. They particularly point to the need of studying the impact of such visualizations on student learning, addressing the challenge of integrating them in realistic learning activities.

References


Concept Maps as Cognitive Visualizations of Writing Assignments

Jorge Villalon1,2 and Rafael A. Calvo1

1School of Electrical & Information Engineering, University of Sydney, Australia // 2Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Chile // jorge.villalon@uai.cl // villalon@ee.usyd.edu.au // rafa@ee.usyd.edu.au

ABSTRACT

Writing assignments are ubiquitous in higher education. Writing develops not only communication skills, but also higher-level cognitive processes that facilitate deep learning. Cognitive visualizations, such as concept maps, can also be used as part of learning activities including as a form of scaffolding, or to trigger reflection by making conceptual understanding visible at different stages of the learning process. We present Concept Map Miner (CMM), a tool that automatically generates Concept Maps from students’ compositions, and discuss its design and implementation, its integration to a writing support environment and its evaluation on a manually annotated corpora of university essays (N=43). Results show that complete CM, with concepts and labeled relationships, are possible and its precision depends the level of summarization (number of concepts) chosen.

Keywords

Concept Map Mining, Concept Map, Automatic generation, Text Mining, Writing

Introduction

University students are expected to develop higher-order cognitive skills such as analysis and synthesis, and also meta-cognitive skills. The first goal is often tackled through writing activities, arguably the task where higher cognitive functions are best developed (Emig, 1977). The second goal of developing meta-cognitive skills can be facilitated by cognitive visualizations (also known as knowledge visualizations) (Jacobson, 2004). Cognitive visualizations are tools that make “thinking” visible, reifying learners’ mental model about domain knowledge onto an explicit graphical device. They make possible the application of “cognitive apprenticeship” approaches like reflection and scaffolding (Collins, Brown, & Holm, 1991). Reflection can be supported by confronting the learner with a different visualization of her own knowledge, while scaffolding can be implemented by contrasting the learners’ visualization to that of an expert.

One cognitive visualization technique is Concept Mapping. Concept Maps (CM) represent a person's understanding of a topic by mapping concepts and their relationships in a hierarchical way, where more general concepts are placed higher in the map and concepts at the same level of generalization are grouped together. There is extensive evidence that drawing a CM requires students to engage in higher cognitive functions (Novak & Gowin, 1984). CMs have typically been used in reading activities to aid students’ comprehension of texts. For instance, ready made CMs may be presented as semantic summaries of texts that students need to comprehend (Hauser, Nuckles, & Renkl, 2006), or students may be asked to construct their own CMs to address specific questions (Chang, Sung, & Chen, 2002).

Our own work presents a novel approach in the educational application of CMs. Here, CMs are embedded in writing (as opposed to reading) activities and are used to summarize the students’ own writing. Unlike in the more typical scenarios of using CMs to support reading, in our work the CMs become approximate representations of students’ current state of knowledge. From the students’ perspective, such CMs can be used to reflect on their own knowledge and also to help students see their writing from a different perspective. From the instructor’s point of view, such CMs can be used as a rapid assessment of students’ conceptual understanding.

One challenge to using CM as a way to support students’ writing, is that it is time consuming and expensive. Automated tools for generating CMs aim to reduce the workload (of the instructor) or redirect it to other learning activities that support the expected outcomes. Attempts to automatically generate CMs have so far been limited to approaches that generate pseudo-CMs where the relationships between concepts are not labeled, and with evaluations that cannot be replicated in a new experimental setting. Our research program has been to use linguistic analysis and statistical techniques to automatically generate complete CMs. We have incrementally stepped towards the automatic extraction of full CMs. First we proposed the idea and the evaluation framework (Villalon & Calvo, 2008), then the first results for a concept extraction algorithm (that did not include relationships) (Villalon & Calvo, 2009). This work required a linguistic and morphological analysis of the benchmarking corpus that was reported (Villalon,
Despite this progress, many challenges remain, the first set being of a technical nature. The automated CMs described in previous work were still significantly less accurate and complete than those built by humans, they do not even have the required relationships between the concepts. The second set of challenges is of a pedagogical nature. Concept maps have so far always been prepared by human readers. The research literature has focused on how they are useful for students to describe (and develop) their conceptual schemas on a certain topic (e.g. on something they have recently read). Since automated CM are only now becoming possible, there is no clear understanding of how they can be used in real learning situations. This new technical opportunity brings the challenge of figuring out the new pedagogical opportunities. This analysis requires a review of learning activities where they can be used, particularly reading and writing where CM can provide a means for developing cognitive and metacognitive skills.

The first contribution of this paper addresses one of the key technical challenges, a technique for automatically producing labeled relationships. The second contribution of the paper is a review of possible learning activities using automated CMs. In particular, next section reviews aspects of the literature on academic writing and concept mapping as learning activities, supporting the pedagogical motivation of the framework. We also review current progress on automated techniques for concept map mining. We show how CMs can be useful to support writing activities, as they are for reading. We follow with a description of a new CMM tool and its integration into Glosser (Calvo & Ellis, 2010) a tool for automated feedback in writing activities. This new tool provides a means for the meta-cognitive skills described earlier by showing the student a different perspective of their own writing. Furthermore we provide an evaluation using a collection of student essays. We finalize with a discussion on the implications of our work.

**Background and Related Work**

**Essay writing**

Essay writing is considered a unique way of learning because it involves en-active (learning "by doing"), iconic (learning "by depiction in an image") and symbolic learning (learning "by restatement in words"). As Emig (1977) explains, writing is "the symbolic transformation of experience through the specific symbol system of verbal language is shaped into an icon (the graphic product) by the en-active hand". This symbolic transformation involves the recall and synthesis of ideas.

A study by Paltridge (2004) found that universities require graduate students from all disciplines to write academically. Another study by Moore and Morton (1999) found that the academic essay represented almost 60% of writing tasks at both the undergraduate and graduate levels. The increased emphasis on writing may reflect educational leaders' recognition of the connection between writing and knowledge production in an information era (Warschauer & Ware, 2006). However, writing academic essays is challenging. It requires much more than good surface writing skills such as producing grammatically correct sentences. As Paltridge (2004) observed, writing academic essays demands deep understanding of “the roles played by the people involved in the production of the texts, and the contexts in which the texts are produced and assessed”.

**Automatic feedback in the writing process**

One way to help students learn to write is by providing feedback on their writing process. In considering writing as a process, Keh (1990) explains that "feedback is the drive which steers the writer through the process of writing on to the product". It is from feedback that writers learn if she "has misled or confused the reader by not supplying enough information, illogical organization, lack of development of ideas”. Quality feedback makes the author's thinking visible and allows the author to reflect on her own work (Lin, Hmelo, Kinzer, & Secules, 1999). Regrettably, meaningful and timely feedback is difficult and expensive to provide, so researchers have been working on automated and semi-automated approaches to feedback. So far, most of the work on automated feedback has been on assessment or written suggestions on features that students should improve on (Beals, 1998; Graesser & McNamara, in press; Thiemeyer & Thiemeyer, 1990; Wade-Stein & Kintsch, 2004; Wiemer-Hastings & Graesser, 2000). Alternative visual representations of their own writing provide a new form of feedback (Calvo & Ellis, 2010).
Concept Maps

Concept Maps (CM) were introduced by Joseph Novak as a way to assess children's understanding of science with graphical tools to organize and represent knowledge (Novak & Gowin, 1984). In a CM, concepts are represented in boxes that are linked by labeled relationships; two related concepts (including their link) form a proposition or semantic unit. Concepts are also arranged hierarchically such that more general concepts are located higher on the map and specific concepts such as examples are located lower. Novak defines a concept as "a perceived regularity in events or objects", or "records of events or objects" designated by a label. A concept by itself does not provide meaning, but when two concepts are connected using linking words or phrases, they form a meaningful proposition.

Since inception, CMs have been widely used and tested as pedagogical tools (c.f. Novak, 2007). A study by Hauser, Nuckles, and Renkl (2006) compared a control group to several groups using CMs in four ways: constructing CMs from scratch, constructing CMs from a list of concepts, constructing CMs from spatially arranged concepts, and finally studying previously built CMs. The results showed that constructing maps from scratch and studying previous constructed ones led to significantly better learning outcomes than the other conditions. A similar study by McClure presented evidence on the validity and reliability of concept maps as an assessment tool (McClure, Sonak, & Suen, 1999). Another study validated the use of concept maps to improve text comprehension and summarization (Chang, Sung, & Chen, 2002). This study compared a control group to three concept mapping approaches: showing an expert generated CM; scaffolded concept mapping (completing a partially-completed expert CM), and constructing CMs from scratch. The results indicated that all approaches to concept mapping improve text comprehension and summarization skills, with the scaffolded concept mapping approach leading to the best outcomes. CM can also be used to scaffold students' reflection. Studies by McAleese show that CMs can be used as mirrors or assistants to the learner. These studies also describe the process of “off-loading” concepts from the mind to the map, and identify several cognitive steps that people take when constructing CMs (McAleese, 1998).

CMs are not the only cognitive visualization technique for documents and text. Recently proposed visual representations including Word Trees (Wattenberg, & Viegas 2008) and Radial Document Visualization (Collins, Carpendale, & Penn 2009) are incorporating relevant semantic information that are improvements over simple statistical visualization like Tag Clouds. Even though such representations are a step forward to a semantic view of a document, they present two limitations when compared to CMs. Word Trees present keywords in the context they appear (typically sentences), however different keywords are not connected to each other as concepts are connected in CMs. Such connections are key to represent meaningful propositions, that give CMs their knowledge richness. Radial Document Visualizations are built in a more semantic way, using connections between concepts, however these connections are limited to “is-a” relationships (hyponymia), our own research has shown that such connections in student essays count for only a small part of the knowledge (Villalon, Calvo, & Montenegro 2010).

Previous studies on automatic concept mapping

A study by Valerio and Leake (2006) analyzed the requirements for generating CMs from text for educational purposes. These requirements can be summarized into the three principles of Educational utility, Simplicity, and Subjectivity (Villalon & Calvo, 2008), elaborated below:

- **Educational utility**: There is scientific evidence for the educational utility of concept mapping following Novak's method. CMs must include concepts, which should be connected by linking words to form propositions. They must also have a topology where more general concepts are placed higher in the map and specific concepts lower. Concepts with the same level of generalization should be placed at the same level (Novak, 2007).

- **Simplicity**: CMs are mainly used for human analysis, giving teachers and students an alternative, structured, and cheaper representation of students' understanding. Novak's definition also indicates that a CM should require no more than 25 concepts to answer its focus question. Since there can be many more ‘concepts’ in a medium length document, CMs (either human made or automatic) can be conceived as a visual summary of the complete document.

- **Subjectivity**: As mentioned earlier, CMs represent both the author's knowledge and her writing skills. In an educational context, the terminology used by the student is also important for assessing the outcome, so CMs should be represented using the terms that the author used in her text. If a student uses a certain word to refer to a particular concept, this choice inevitably reflects her vocabulary level, and hence a concept map should retain
this information. This means that poorly written essays should be able to produce a CM with meaningful propositions that should also reflect the author’s writing skills limitations (like spelling and grammar errors).

Several studies have focused on the automatic generation of CMs from text. These studies expose challenges at two levels: An inconsistent definition of CMs, and a lack of an evaluation framework. The first problem occurs when authors report on the automatic creation of CMs, but are actually creating a different type of knowledge representation with its unique characteristics, which do not follow Novak's definition. The second problem is that each study uses a different method to evaluate its contribution, making it very hard to compare them. For a detailed description of previous attempts see (Villalon & Calvo, 2008).

Three studies described the construction of CMs from documents to facilitate the visualization of domain knowledge (Chen et al., 2008; Clariana & Koul, 2004). These studies create maps with concepts that are connected by unlabeled relationships and generate propositions which do not conform to Novak's (2007) method. Three other studies addressed these limitations, creating CMs which include some type of labeled relationships. Oliveira, Pereira, and Cardoso (2001) presented a system called TextStorm which generates 'raw CMs' from texts for students to elaborate further. These authors used a rule-based approach that can analyze only affirmative and declarative sentences. The CM quality was evaluated by manually assessing propositions generated from a set of articles, essays and book chapters. The second study, by Valerio and Leake (2006) also proposed the idea of using automatically generated CMs as an educational tool to 'jump-start' the CM construction. In this study CMs were evaluated in an application based fashion. The generated concepts were used to link new documents to improve precision and recall in Information Retrieval. The approach does not evaluate relationships or the CM's fidelity to the text source. A study by Zouaq and Nkambou (2008) presented a system named TEXCOMON that creates CMs as part of an Ontology Learning from Text (OLT) process. Using linguistic analysis they generate a CM from a corpus and use it to infer an Ontology that represents the corpus knowledge. They evaluated the quality of the generated ontology by comparing it with an ontology generated using state-of-the-art OLT tools.

The Concept Map Miner (CMM)

Within our research program for providing automated feedback for writing activities we have shown progress in writing support tools allow that feedback to be provided at any stage of the writing process (Calvo et. al., 2011). This is a significant improvement over previous efforts that focused on providing feedback on the final product that students submit. We have also showed new approaches to help students reflect on their writing (Calvo & Ellis, 2010) and how students understand the use of these new tools. Concept map visualization fall within this design strategy where the tools help students reflect about their own writing, helping them develop metacognitive skills.

Mathematically, a CM is defined as a triplet CM={C, R, T} where C is a set of concepts, R a set of relationships between concepts, and T is the map's topology or spatial distribution of the concepts. The CM mining process can be expressed as the identification of a concept map CM from a document D. This process has three steps: Concept Identification (CI); Relationship Identification (RI); and Summarization, which reduces the size of the CM by keeping only the most relevant concepts. As both concepts and linking words come from a single document, D itself defines all potential words (or phrases) that could become part of the CM. This can be formalized by defining a document as a duplet $D=\{C_d, R_d\}$ where $C_d$ corresponds to all the concepts, and $R_d$ corresponds to all the relationships that explicitly appear in the text. The whole process is summarized in Figure 1.
The CMM tool was implemented as part of a project to develop TM algorithms for educational purposes (Villalon, Kearney, Calvo, & Reimman, 2008; Calvo & Ellis, 2010; Calvo, O’Rourke, et al., 2011). The automated language processing functionalities have been incorporated into a Text Mining Library (TML) that combines linguistic analysis and statistical techniques to support semantic analysis of student essays. Conceptually TML performs ‘operations’ on corpora. An example of an operation would be to calculate the semantic distance between each document in the corpus formed by all documents containing a particular word.

The first step in the CMM process is to identify concepts and relationships. We have identified that compound nouns worked well for concept identification, information that was extracted exploiting the grammatical tree of each sentence (Villalon & Calvo, 2009). For relationship identification, grammar trees did not carry enough semantic information. A new semantic layer was obtained using typed dependencies, which define a dependency tree for each sentence. Each dependency tree is then transformed into a terminological map, equivalent to the one shown in Figure 2. Terminological map rules are then applied to the map to transform it into a reduced version with no grammatical dependencies. For example, several rules are used to form compound nouns required for concept identification, like for the triplet language – amod – artificial, which indicates that artificial is an adverbial modifier of language, therefore it corresponds to the compound noun ‘artificial language’.

![Figure 2. Terminological map of a sentence](image)

![Figure 3. Extracted CM for a sample essay using CMM](image)
Both Concept and Relationship Identification are implemented as operations in TML, using the extracted terminological maps with all terminological map rules applied to obtain a reduced map. The new Concept Identification operation extracts all vertices in the map corresponding to nouns or compound nouns, instead of extracting them from a grammar tree. The Relationship Identification (RI) operation uses the terminological map and a set of concepts as input (the results of a Concept Identification operation). RI has been implemented by finding the shortest path between each pair of concepts, using Dijkstra's algorithm in the terminological map. Redundant connections were avoided by discarding paths containing concepts, and paths that were equivalent but in an opposite direction. Remaining paths are used to identify relationships, and the relationship labels (linking words) are obtained as the concatenation of the words in the path. Some dependencies such as conjunctions and prepositions also include words, which were cleaned and then included. The final summarization operation was implemented using Latent Semantic Analysis (LSA), as suggested in our previous work. LSA is a statistical technique used to find patterns on big corpora (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), that was found to be useful to identify the relative importance of terms within a document (Villalon & Calvo, 2009). A complete CM for a sample essay is shown in Figure 3.

**Relationship Extraction and CMM**

The quality of automatic concept maps depends on the quality of the knowledge extraction techniques used. Measuring the performance of these techniques first requires that a group of human annotators build a ‘gold standard’ corpus with annotations that have high inter-coder agreement and are compared to those extracted automatically. Identifying knowledge in text is a subjective task which involves the judgment of humans in the face of ambiguity, such as when a word has many meanings, whether a sentence should be part of a summary, or if a phrase had a particular intention. Under the assumption that humans can reliably make such interpretations, subjectivity is tackled by using corpora annotated by two or more human coders who are required to identify the relevant pieces of knowledge in a corpus. Inter-coder reliability is the upper bound for accuracy for any automatic algorithm trying to identify the same set of knowledge (Artstein & Poesio, 2008).

**Data**

A set of essays (N=43) collected as a writing proficiency diagnostic activity for first year-university students were used to create the benchmarking corpus. In writing the essays, students first read three short papers on the topic of *English as a global language*, and then answered two questions: Is English becoming a global language? Is this a positive development? The essays had an average length of 468 words; the corpus had a total of 18,431 words.

A first version of the benchmarking corpus was annotated by two coders using a web based annotation tool described in (Villalon & Calvo, 2008), and used to measure the accuracy of our first version of the Concept Extraction algorithm proposed in (Villalon & Calvo, 2009), we refer to this corpus as the first annotation trial. Even though this corpus was useful for extracting concepts, it presented a very low inter-human agreement for relationships, the main problem found was that coders created relationships that were not explicitly present in the essay, but were an interpretation of several propositions. As each coder interpreted the text based on her own knowledge, the already subjective task of summarizing an essay using a CM became even more subjective, resulting in poor inter-rater agreement. To improve this, the protocol was modified, as described below. The tool was also modified to verify this at addition time. Two new human coders using the modified tool and the new protocol performed a second trial that improved inter-rater agreement. Quantitative data on this improvement is reported in the results section. The process averaged 23 minutes per CM, and produced a total of 440 concepts and 258 linking words per coder.

**Annotation Method**

To annotate the benchmarking corpus for the evaluation of the CMM, the human coders need to follow the same protocol (Artstein & Poesio, 2008). We followed Novak’s (2007) method for constructing good and generic CMs that reflect the understanding contained in a document. The first step consists on stating a good focus question that will guide following steps. Then, a list of the most relevant concepts is created; concepts in this list must be ranked from the more general/inclusive to the more specific. Novak refers to this list as "the parking lot", from where
concepts are taken and put in the concept map one by one, starting from the most general ones, and linking each one to previous concepts when needed, forming good propositions. He explains that previously chosen concepts can be discarded if they do not provide new information to the map (Novak, 2007).

The annotation protocol can be summarized as follows:
- Identify the focus question that the essay is trying to answer.
- Make a list of concepts that are the most relevant in the document using words from the document, including spelling and grammar errors.
- Rank the concepts from the most general to the most particular according to what the document explains.
- Begin the concept map with 1 to 4 of the most general concepts.
- Choose explicit linking words to relate the concepts; the words must have appeared in the document and must relate the two concepts. To ensure that the concepts are related, both must appear in the same paragraph.
- Add concepts and relationships until the knowledge expressed in the document is accurately represented in the concept map.
- Reposition the concepts if necessary, to reflect the different levels of generalization of the concepts expressed in the document, in the concept map.

An analysis of the resultant Gold Standard was performed to validate that the CMs were compliant with their educational requirements. For Educational utility, all CMs had concepts and labeled relationships and only very few of them had isolated concepts, human annotators reliably reported that such cases represented isolated ideas in the essay. For simplicity, all CMs were of a considerable smaller size than their corresponding essay, even though the number of concepts used correlated positively with the length of the essay, the number of relationships (inter-connected concepts) did not correlate with the document extension. Finally, for subjectivity, the essay grades showed no correlation with the CMs size, measured in concepts and relationships. This result showed that poorly written essays can be represented by a CM even if they reflect poor writing skills. A more detailed discussion on how humans summarize text using CMs can be found in (Villalon, Calvo, & Montenegro 2010).

Comparative Measures for CMs

Two quantitative measures that can be applied to compare CMs have been proposed in the literature: The Kappa statistic that measures the extent to which two coders agree above chance, and Information Retrieval for Ontology Learning from Text (OLT) measures (Dellschaft & Staab, 2006). A measure first proposed in the information retrieval literature, corresponds to the lexical layer, which is the identification of 'concepts' in the text. These methods are evaluated using the Lexical term Precision (LP). The hierarchical layer corresponds to relationships between concepts. Several measures have been proposed and discussed for this, and the current state of the art corresponds to Taxonomic Overlap Precision (TP) based on the common semantic cotopy proposed by Dellschaft and Staab (2006). LP measures how well the learned lexical terms (concepts for CMM) cover the target domain and TP measures how well the concepts are arranged in the hierarchy defined by an ontology (Dellschaft & Staab, 2006). An important quality of such measures is that they are influenced only by one dimension, and therefore affected by one type of error only. As Dellschaft and Staab point out "if one uses measures for evaluating the lexical term layer of an ontology (e.g. the lexical precision and recall) and one also wants to evaluate the quality of the learned concept hierarchy (e.g. with the taxonomic overlap), then a dependency between those measures should be avoided." (Dellschaft & Staab, 2006) In order to achieve this, the common semantic cotopy must be used to evaluate TP only for those concepts that are common to both maps.

Results

The results of the inter-human agreement show that the modified annotation protocol improved its reliability. Inter-human agreement was measured for the two annotation trials. In the first trial, making a list of the most important concepts, showed an average of 62% agreement for LP and 27% for TP. This shows that humans tend to agree when choosing the most relevant concepts in a document; however, they tend to disagree about linking concepts if given too much freedom. The second annotation trial showed an average of 77% agreement for LP and 85% for TP, indicating a small increase in agreement on concepts and a large increase in agreement on relationships. This increase cannot be caused by chance, because the number of possible linking words for two concepts are only
restricted to all words (not just a category like nouns or verbs) appearing in the same paragraph. Therefore, the possibility of connecting two concepts in the same way by chance is approximately 0.

The automatically generated CMs were compared to the human generated gold standard, showing good results that are affected by the summarization rate. As summary visualizations, an automatic CM should contain only the 'most important' information in an essay. The main characteristic of summaries is their 'size', which is defined as a percentage of the total text units (units can be any subpart of the text, such as sentences, words, terms, paragraphs). For example, a CM reduced to 10% will contain only 10% of the total concepts identified and all relationships between those concepts. In the Concept Map Miner tool, the strategy for summarization was to apply LSA to each essay, forming single document semantic spaces that sort topics within the essay by their relative importance. Figure 4 shows the effect of summarization on agreement measures.

A high lower bound for LP of 50% validates LSA as a good strategy for summarization. LP grows rapidly reaching 80% with CMs of less than half their original size. Summarization causes a reduction of up to 20% on TP, this results are good because improvements on relationship extraction should improve results even at strong summarization ratios. Finally, LSA does not use the information about the CM, making connected concepts to be removed if they do not provide information, and causing all its relationships to disappear. An approach that would consider both the information provided and the original CM and its connections while summarizing, may provide better results.

![Figure 4. Lexical and Taxonomical Precision versus Summary size](image)

Integration of CMM as Writing Support Tool

Cognitive Visualizations provide quality feedback because they make the author’s thinking visible, making explicit the mental model learners are using. CVs can also function as metaconceptual scaffolds, allowing learners to develop an awareness of their own mental representations and inferential processes (Jacobson, 2004). Within the writing process, providing automatic feedback in the form of CVs allows learners to reflect on their own work and their own mental models that guided its construction, facilitating the development of metacognitive skills. One way to better support the writing process is by improving feedback’s availability through automatic feedback, in this way the learner can obtain such feedback as many times as she requires, therefore allowing metacognitive skills to develop through the writing process. This is a common design principle in the set of tools (a.k.a. Glosser) that our research group has been developing by using TM techniques to provide automatic feedback to students (Villalon, Kearney, Calvo, & Reimman, 2008). By integrating CMM into our Glosser tool (Calvo and Ellis, 2010) and using cloud computing technologies (Calvo et.al., 2011) students can access feedback on the current state of their assignment at
any point in time. A common learning activity using CMM to support essay writing would be: A teacher uploads readings to the University's Learning Management System and creates an assignment, students then write their essays in the Google Docs tools, with the documents produced by the assignment manager in iWrite. Before (or after) the deadline students can use CMM or any of the other Glosser feedback tools. This is a significant improvement over having to wait a few weeks to receive feedback from the instructors.

CMM is designed to be used in real writing activities aimed at helping a student learn about a topic, rather than focused on communication as a standalone skill. In many writing activities students are asked to write an essay about a topic. This activity can take from a few hours to a full semester. Using CMM, at any point in time, the student can see a CM of the composition and evaluate if those concepts and relationships were what he expected (and what is expected from him). CMM can also be used by the instructors, as other studies have shown that visualizations of the essays can improve the time and reliability of human assessors (O'Rourke, Calvo, McNamara, 2011).

We have integrated CMM into the Glosser environment for automatic feedback. Glosser’s designed is aimed at provoking reflection in students during the writing process and details on its pedagogical design are discussed in (Villalon, Kearney, Calvo, & Reimman, 2008). The feedback is in the form of descriptive information about different aspects of the document. For example, by analyzing the words contained in each paragraph, Glosser can measure how close two adjoining paragraphs are. If the paragraphs are too far this can be a sign of a lack of flow or what is called lexical in cohesiveness and Glosser flags a small warning sign. Glosser provides feedback on several aspects of the writing: structure, coherence, topics, keywords. Each of these areas is identified by a tab on the homepage of the web application as shown in Figure 5. The Questions section contains triggers to prompt student reflection on a particular set of features of the composition, and can be customized for each course or activity. The instructions section describes how to interpret the feedback, which is provided as 'gloss' on the lower half of the page. The feedback is an alternate type of visual or textual representation of features that have been automatically sourced from
the text. The history of versions provides a record of the writing process, which includes information on the time of each revision and the author responsible for it.

Another way to improve the quality of feedback in the writing process is by improving the quality of the CVs used as feedback, and CMs can provide such quality. As Jacobson argues: “One technique for doing this (making thinking visible) is the use of concept maps to represent knowledge” (Jacobson, 2004). However, until now it was impossible to automatically extract such a CV from students’ text. Using the CM to visualize their own mental models, students can compare them to a model map provided by the teacher (as in cognitive apprenticeship) or revise their writing so the CMM tool can clearly identify the parts of the model it could not automatically extract. Reflection for revising is scaffolded in Glosser by providing questions for the student to reflect when analyzing the CM.

The CMM integration consisted on a new tab labeled “Concept Map” that shows the generated CM, and presents the questions that can be customized by each instructor:

- Do you think the most relevant concepts you covered in the essay are present in the map?
- Could you improve the Concept Map by adding or removing concepts and relationships?
- Do you think someone could understand your argument only by analyzing this map?

Conclusion

Motivated by the pedagogical value of Concept Maps as cognitive visualizations, this paper presents a new Concept Map Mining (CMM) tool, and evaluates it using a collection of human annotated essays written by undergraduate students. The results show that the automatic generation of CMs from documents is feasible, despite the complexities of noisy data such as student-produced text. Humans tend to agree when summarizing a document using a CM (inter-coder agreement of 77% for Lexical Precision and 85% for Taxonomical Precision). Available OLT tools and the CMM reliably identified concepts, averaging 94% for LP with human coders.

The tool has been integrated into an e-learning environment, and future work involves designing pedagogical activities in which the impact on student learning will be assessed. As the motivation of the CMM is to support writing activities, it was also integrated into a tool for enhanced feedback on writing activities called Glosser (Villalon, Kearney, Calvo, & Reimman, 2008; Calvo and Ellis 2010). The tool supports writing activities by scaffolding authors' reflection during the process of writing, to encourage them to revise their work. At any time during their writing authors can gloss their documents by clicking a link that will take them to the Glosser website. Glosser can then present several issues related to the quality of the essay with provoking questions on how to assess them. Automatically generated features from the text are also presented for the student to guide and facilitate the answering of the questions.

Acknowledgment

The authors would like to thank Stephen O'Rourke for his contribution to this project and Anindito Aditomo for feedback on the manuscript. This project has been funded by an Australian Research Discovery Grant DP0665064, a Norman I Prize and a Chilean Bicentennial Scholarship.

References


**Knowledge Visualization for Self-Regulated Learning**

**Minhong Wang**, **Jun Peng**, **Bo Cheng**, **Hance Zhou** and **Jie Liu**

1Faculty of Education, The University of Hong Kong, Hong Kong // 2College of Science, Nanjing University of Aeronautics and Astronautics, China // magwang@hku.hk // woaipjma@hku.hk // bob88708@hku.hk // zhouhance@nuaa.edu.cn // jieliu@hkusa.hku.hk

*Corresponding author*

**ABSTRACT**

The Web allows self-regulated learning through interaction with large amounts of learning resources. While enjoying the flexibility of learning, learners may suffer from cognitive overload and conceptual and navigational disorientation when faced with various information resources under disparate topics and complex knowledge structures. This study proposed a knowledge visualization (KV) approach to this problem in an online course. The investigation involved the design, development, and evaluation of an enhanced learning system for the course using the proposed approach. The focus was on visualization of domain knowledge structure and integrating the structure with curriculum design, learning resources, learning assessment, intellectual process, and social learning. Survey and interviews with students demonstrated high user satisfaction and acceptance with the developed system and its functions for KV. These findings lay the foundation for further exploration with the system to determine its impact on reducing cognitive load and improving the learning process.

**Keywords**

Knowledge Visualization, Online Learning, Knowledge Structure, Self-Regulation, E-Learning

**Introduction**

Due to its flexibility in delivery and just-time access, e-learning has been widely adopted in recent years. In e-learning applications, learners are encouraged to learn through interacting with a wide range of resources to acquire and build their knowledge. While such a resource-abundant and self-regulated learning environment allows learners a great deal of freedom and flexibility in searching for, selecting, and assembling information, learners may suffer from cognitive overload and conceptual and navigational disorientation when faced with massive information online (Tergan, 2005; Kayama & Okamoto, 2001; Miller & Miller, 1999). The challenge is even greater when learning contents are scattered under disparate topics and complex knowledge structures. When faced with this problem, many learners are unable to figure out features and meaningful patterns of various kinds of information, and are easily hampered by limited working memory. This is mainly because novices lack sufficient knowledge and a deep understanding of the subject domain, which is crucial to organizing information and knowledge for retention in long-term memory. Also, traditional education breaks wholes into parts, and focuses separately on each part, and learners are often unable to create the big picture before all the parts are presented. As a result, most online learners, especially novices, become “lost-in-hyperspace”.

This study aims to improve the design of current e-learning systems by dealing with the aforesaid problem. To facilitate cognitive processing and self-regulated learning, learners should be supported with appropriate learning strategies, among which cognitive and meta-cognitive strategies have been well identified (Bransford, 2000; Zimmerman, 2000; Winne, 2001). Learners are helped in their independent learning if they have conceptual knowledge, and learners can become more independent if they have awareness of their own knowledge and ability to understand, control, and manipulate individual learning processes. While these strategies have been found to be effective, few studies have examined how these strategies can be implemented in instructional design, especially in online learning environments. While learning theories or strategies offer guidelines of improving the design of current e-learning systems, it is far more difficult and additional effort is needed to explore effective instructional methods (Reigeluth, 1999).

This study investigates a knowledge visualization (KV) approach to support resource-abundant and self-regulated online learning, which consists of three components. First, an explicit representation of conceptual knowledge structure is constructed by capturing key knowledge concepts and their relationships in a visual format. This visualized knowledge structure serves as a cognitive roadmap to facilitate the knowledge construction and high level thinking of online learners. Second, abstract concepts are connected with concrete contents by linking knowledge concepts with learning resources. In this way, information processing and knowledge construction, the two key
aspects of the learning process, are well integrated. Learners can easily navigate throughout the resource-abundant, non-linear knowledge space aided by the visualized cognitive roadmap. Third, meta-cognitive learning support is provided for learners to regulate and plan their learning process. Assessment materials associated with knowledge concepts are provided for self-evaluation of learning outcomes in granular knowledge components, from which the system generates feedback and guides individuals throughout their learning process.

To implement the proposed approach, an online learning system was developed using computer and Web-based technologies. The system has been designed to help learners transcend the limitations of their minds, not only in cognitive processing, but also in high level thinking and knowledge construction. In doing so, computers are used as electronic pools for reflecting human cognitive processes through visual representations on the screen. These visual representations provide more effective use of learners’ mental effort by amplifying, extending, and enhancing human cognitive functions, and engaging learners in representing, manipulating, and reflecting on what they know.

Compared with other related work, this study is unique in the following aspects. First, while traditional education breaks wholes into parts and focuses separately on each part, this study aims to help learners see the “whole” before they are able to make sense of the parts. Second, instead of asking learners to construct knowledge maps by themselves, this study utilizes expert knowledge structure to help novices build up their thinking and understanding on a solid foundation. This may reduce novices’ cognitive overload in advanced thinking. Third, instead of using visualized knowledge structure as an isolated instructional instrument, this study uses it as infrastructure and integrates it with curriculum design, learning resources, learning assessment, intellectual processes, and social learning.

**Proposed Knowledge Visualization (KV) Approach**

This section presents theoretical dimensions of the proposed approach.

**Visualization of Knowledge Structure**

In facilitating learners’ cognitive processing and retaining knowledge in long-term memory, clustering or chunking (i.e., organizing disparate pieces of information into meaningful units) is regarded as a pervasive approach (Simon, 1974; Bransford, 2000). According to psychology theories, knowledge in memory is organized semantically in networks, built piece by piece with small units of interacting concepts and propositional frameworks. These mental semantic networks represent a cognitive structure, which can be used as a learning tool for constructive learning processes (Jonassen, 2000). More importantly, the cognitive structure should be represented in an external format with explicit description. This is because visual methods help externalize and elicit the abstract structure of knowledge (Jacobson & Archodidou, 2000), and human brains have rapid processing capabilities to acquire and retain visual images (Paige & Simon, 1966). Computer-based technologies help further by making it easy for learners to construct, recall, and modify visual representations, and keep them for a long period of time (Jonassen, 2000; Novak & Cañas, 2008).

The KV approach proposed in this study is to incorporate visualized representations of domain knowledge structure into e-learning systems. Relevant functions are developed for creation, storage, display, and revision of knowledge maps. Rather than memorizing the content, learners can use knowledge maps to identify important concepts and their relationships, and generate semantic networks for review and reflection. Moreover, a knowledge map displays intellectual processes involved in the acquisition and construction of knowledge. These become the basis for systemic inquiry, knowledge construction, and high level thinking (Wang et al., 2010).

**Integration of Information Processing and Knowledge Construction**

Information processing and knowledge construction are closely intertwined in the learning process. Learners need to access information to acquire content knowledge and formulate hypotheses (Jonassen, 1999). Knowledge is constructed through meaningful learning, which takes place when learners deliberately seek to relate (new) information to, and incorporate it into, relevant knowledge that he/she has already possessed (Mayer, 2001).
Objectivism and constructivism offer different perspectives on the learning process, and provide complementary guidelines on how learning should be engendered (Jonassen, 1999; Miller & Miller, 1999). Information processing theory, which regards a human as an information processor or a “mind as computer” (Newell & Simon, 1972), is based on the objectivist paradigm, which assumes that knowledge has an objective and separate existence whose attributes, relationships, and structure can be known, and that learning is transmission of knowledge from teachers to learners (Reynolds et al., 1996). The constructivist paradigm, on the other hand, purports that knowledge is not independent of the learner, but is internally constructed by the learner as a way of making meaning of exercises (Jonassen, 1999). This study intends to integrate and facilitate both objective information processing and subjective knowledge construction in e-learning by linking knowledge concepts in visualized maps with learning resources.

Facilitation of Self-Regulated Learning

Advanced learning acquires over years of experiences and derives from activities of thinking, action, and reflection. Experts have acquired a great deal of well-organized content knowledge, and their organization reflects a deep understanding of the subject matter (Bransford, 2000). Although peer models have been used to guide self-regulated learning, the creation of well-developed and stable cognitive structures for scaffolding advanced learning is noted as a primary instructional goal (Zimmerman, 2000; Reigeluth, 1999). The recognition of expert knowledge has been reflected in both objectivist and constructivist learning theories. Information processing, which is based on objectivist learning theory, requires effective and efficient processing of information and indicates that experts’ knowledge structures help learners acquire information accurately. At the same time, constructivism suggests that guidance and strategies (e.g., modeling, coaching, and scaffolding) from experts provide the necessary support for learners to construct knowledge (Miller & Miller, 1999).

This study utilizes expert knowledge structures for guiding and scaffolding novices’ understanding, thinking, and inquiry in their self-regulated learning. Conceptual understanding of a domain is often not fully expressed in books or learning materials, and knowledge maps can be used to articulate and manipulate such tacit knowledge more effectively. Using the expert knowledge map as the foundation may reduce the chance of misconception and faulty ideas. Although highly structured graphs may seem constraining at times, these templates are good starting points for novices, who have trouble organizing their understanding and are confused in their self-regulated learning (Hyerle, 2000).

In addition to facilitating learners’ cognitive processes, knowledge maps provide meta-cognitive support. As mentioned, learners can become more independent if they are aware of their learning process and have the ability to regulate it. Visual representations are forms of metacognition that graphically display the thinking process (Costa & Garmston, 2002). Knowledge maps display intellectual processes by representing sequences, alternatives, branches, choice points, and pathways that involved in the acquisition and construction of knowledge (Wang et al., in press). To utilize this metacognitive feature, additional functions based on knowledge maps were developed in this study to help learners plan and oversee their learning process. In doing so, assessment materials were collected and associated with knowledge concepts for evaluation of learning outcomes in granular knowledge components, with feedback to learners for correct answers and detailed explanations. At the same time, the system can monitor individual learning progress, based on which learning guidance is provided to individuals such as what to learn in the next step, further effort required for a specific knowledge concept, reminder of prerequisite knowledge to learn before moving on, etc. Individual learning progress can also be reflected in the knowledge map to indicate the knowledge that has been learnt, ready to be learnt, or not ready to be learnt.

Support of Social Learning

To support self-regulated learning, learners are encouraged to participate in social communication, discussion, and sharing. The knowledge structure constructed in this study can also be used as the index or model to organize discussion messages and shared learning resources, with a view to facilitating and steering social communication and knowledge sharing in the social learning community (Wang, in press).
Related Work

To deal with the cognitive load problem, researchers on cognitive load theory have recommended the use of schema to categorize multiple information elements into a single entity (Sweller, 1994). However, with their focus on reducing the number of information elements by organizing them around entities such as categories or examples, these studies did not examine how different types of entities can be identified and constructed, and how learning can be supported by providing a cognitive structure for high level thinking and understanding. Other related research has looked into the cognitive load problem from the multimedia learning perspective (Mayer, 2001). With a focus on simultaneous processing of visual and audio information, these studies recommended that the cognitive load of both channels be managed, and aligned or synchronized for a coherent multimedia presentation.

In relation to KV, concept mapping has been used in education for more than thirty years. It is used mainly for students to demonstrate their understanding and ideas and for teachers to assess students’ understanding. Traditional ways of constructing concept maps used paper-and-pencil, and computer-based graphical tools were later developed for expressing knowledge comprising concepts connected by arcs. More recently, the CmapTools software was developed particularly for a Web-based environment, whereby users can construct, modify, and publish their concept maps on the Web (Novak & Cañas, 2008). Although concept mapping has been found to be effective in constructing, communicating, negotiating, and assessing understanding (Tsai et al., 2001; Novak & Cañas, 2008), self-constructed concept maps pose high cognitive demands on learners’ ability to analyze and externalize knowledge. Teacher and students need to undergo extensive training to learn how to integrate multiple forms of thinking into a unified, complex weave of interrelated concepts (Hyerle, 2000; Tergan, 2005). This may make the problem of cognitive overload even greater for novices.

In addition to articulation and clarification of understanding, concept maps have been used for curriculum design and planning (Tergan, 2005; Novak & Cañas, 2008). Coffey (2005) described a software program that provided information and knowledge visualization capabilities to foster orientation and learning. Rittershofer (2005) proposed the use of topic maps to support text-based and graphical navigation through huge amounts of information. In these studies, conceptual structures were used to reflect key concepts and principles to be taught in a course, and suggest optimal sequencing of instructional materials. However, these studies have focused on technical implementation, with minimum pedagogical investigation and empirical report on the use and impact of the solutions in learning experience and performance.

Other studies have used concept maps in developing personalized e-learning or intelligent tutoring systems. Hwang (2003) developed a testing and diagnostic system for providing personalized feedback to learners according to test results and the concept model of the subject matter. The relationships between the subject concepts and test items were set up for analyzing learning problems. Zhuge & Li (2006) built up an individualized online learning system, in which tailor-made learning materials and learning instructions were provided based on course structure, and learners’ background, preference, and learning status. The course structure was separated from and linked with learning materials for flexible reuse of learning materials for individual demand. The main purpose of these studies was to provide personalized learning content and learning instructions based on course structures and learner profiles. With the focus on the reasoning algorithms, concept maps in these studies were hidden inside the system to support reasoning rather than to present knowledge structures to learners. With these developed systems, learners were facilitated to access relevant learning resources instead of developing systemic thinking and understanding based on knowledge structures.

System Development

To demonstrate the effectiveness of the proposed approach, an online learning platform “JAVA E-Teacher” was developed in this study. An entry-level Java-programming course for novices was developed with the system. The course was designed according to the syllabus specified by National Computer Rank Examination of China (NCRE) (http://sk.neea.edu.cn/jsjdj/index.jsp), and the Chinese language was used for the targeted learners. NCRE is an examination and certification system administered by the Ministry of Education of China for evaluation of knowledge and skills in computer operations, computer programming, and network administration. With the increased demand for computer professionals in various fields, NCRE has been well received in China, as evidenced by 28,700,000 participants in various exams from 1994 to 2008, of whom 10,730,000 participants achieved relevant
certificates (http://sk.neea.edu.cn/jsjdj/index.jsp). Java is a popular programming language used by most industry segments for various computing systems; there has been a high demand for Java programmers in recent years. However, most universities of China only deliver C or C++ instead of Java as a regular programming course, and many students have to learn Java by themselves. In view of this situation, a pure online Java course was developed in this study for easy delivery and flexible access by learners. The project is currently for research purposes only, with the aim to investigate how e-learning design can be enhanced through KV for resource-abundant and self-regulated learning.

The knowledge structure of the course was built up by inviting two experts in Java Programming, one with intensive teaching experience and the other with more practical skills and experience. The experts defined the knowledge structure according to the syllabus of NCRE in addition to a number of textbooks and references. It was found that it was difficult and inappropriate to include all the knowledge concepts within one map. The cognitive theory suggests that only about seven items or chunks of information can be held in working memory simultaneously (Miller, 1956). Based on this concern and the nature of the course, the course was first broken down into nine chunks/units: Java Basics, Object-Oriented Programming Fundamentals, Data Types and Operations, Work with Objects, Control Statements, Classes and Methods, Java Applets and Web Applications, Input/Output Techniques, and Graphical User Interface. Each unit contained more or less ten key concepts. In this way, learners might have manageable and measurable milestones and might not feel so overwhelmed. Figure 1 describes a visual representation of the knowledge structure for the Unit “Classes and Methods”, in which eight knowledge concepts are specified, together with their relationships such as composite, prerequisite, and confusing mix. In the following sections, this unit is used as an example for further description of the system.

![Figure 1. A visual representation of the knowledge structure](image)

To implement the knowledge structure in machine language, Protégé, a free open-source ontology editor developed by Stanford Medical Informatics at Stanford University, was used. Using Protégé, knowledge structures can be easily edited and modified via graphical interfaces. Based on the defined expert knowledge structure, a set of visualized knowledge maps were built up. Figure 2 shows some screenshots of the knowledge map for the Unit “Classes and Methods”. The knowledge map is scalable, interactive, and individualizable. Learners can zoom into/out of the map by clicking “+” and “-” around knowledge nodes. Learners can also drag the nodes to adjust the display of the map based on personal preference. During their learning process, learners may receive updates and guidance generated by the system, reflected in the changed color of knowledge nodes, and pop-up messages. A knowledge node displayed in green, blue or grey indicates that this knowledge has been learnt or accessed, ready for learning, or not ready for learning due to some reason (e.g., the prerequisite knowledge is not yet acquired), respectively. The pop-up messages provide learning guidance, e.g., what to learn in the next step.

Moreover, knowledge maps are linked with relevant learning resources, assessment materials, and Questions & Answers (Q&A) discussion items. Learners can click a knowledge node in the map to access associated lecture slides, audio presentations, reading materials, self-tests, and Q&A discussions. The interface shown in Figure 3 works as a nested learning platform for learners to acquire specific knowledge through reading, testing, inquiring and discussing; or to return to knowledge maps to figure out the position of specific knowledge in the domain, or navigate to other concepts.
The assessment questions (most in multiple-choice format) are collected from relevant test databases and references. The questions are reorganized and associated with knowledge concepts to assess learning outcomes in a granular and flexible way. Once a knowledge concept is selected for a test, the test paper will be generated by the system via randomly selecting corresponding questions from the database. Learners may also select to take a test of a unit, for which the system will create a test paper covering all knowledge concepts under the unit. The system will avoid repeated questions when learners take several tests for the same topic. Based on the test results, the system will provide feedback for each question. For a test of a unit, the system will generate a performance report on each knowledge concept in the unit (see Figure 4).
Social learning is also supported in the developed system. As shown in Figure 5, the system provides a Q&A discussion forum for learners to post and respond to questions and inquiries. Each question is linked with one or more knowledge concepts by the initial poster. Questions will be automatically put into three groups, namely, “unresolved questions”, “resolved questions”, and “difficult questions”, according to their status. A posted question is identified as resolved once the initial poster is satisfied with the solution and clicks a button to close the discussion. Questions that have not been resolved within a certain period of time will be identified as difficult questions, for the attention of online tutors to provide necessary support. Moreover, an online chat tool is provided for learners to consult with tutors. Learners are also encouraged to share additional learning resources onto the system. The shared resources can
be linked with one or more knowledge concepts by the contributor, and their appearance on the platform is subject to approval by tutors.

**Evaluation**

To evaluate the proposed approach, a series of experiments and analyses were conducted. At the initial stage of the project, the evaluation was on learners’ perceptions and reactions towards the system and its KV related functions, as reported in this section. The assumption is that unless the proposed KV approach is properly implemented to the extent that learners find it acceptable and satisfactory, further explorations into the effect of the approach on learning may not produce reliable and meaningful results. Learner reactions have been found to have a much larger impact on learning in technology-mediated learning environments (Sitzmann et al., 2008).

**Evaluation Design**

Twenty students from a university in mainland China were recruited as volunteer participants in using the JAVA e-Teacher system. They were final year Bachelor students in the Program of International Business and Trade, with little or no JAVA programming knowledge before participating in this study. Two online tutors (one full time, another part-time) were recruited to provide necessary learning support via the Q&A discussion forum and the online chat tool. It took eight weeks for the participants to complete the online course.

To evaluate the system, a questionnaire survey and an interview were implemented to collect the data on students’ perceptions and reactions towards the overall system and its seven functions developed based on visualized knowledge structure. The functions were: knowledge maps for conceptual and content learning, self-tests around knowledge concepts, guidance on learning navigation within maps, discussion driven by knowledge maps and concepts, upload of Q&A messages with links to knowledge concepts, presentation of Q&A items around knowledge concepts, and online chat and consultation with tutors for conceptual and content understanding. The evaluation was designed to examine students’ responses to five constructs: perceived ease of use, perceived usefulness, perceived satisfaction with the system, attitudes to online learning, and intention to use online learning systems.

**Table 1. Background information of subjects**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Time of using Internet (h/day)</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>7</td>
<td>35.00%</td>
<td>0-1 hour</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>65.00%</td>
<td>1-3 hours</td>
<td>4</td>
<td>20.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3-5 hours</td>
<td>9</td>
<td>45.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5-7 hours</td>
<td>6</td>
<td>30.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>over 7 hours</td>
<td>1</td>
<td>5.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Frequency</th>
<th>Percentage</th>
<th>Time of using Internet (h/day)</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24</td>
<td>19</td>
<td>95%</td>
<td>0-1 hour</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>25-30</td>
<td>1</td>
<td>5%</td>
<td>1-3 hours</td>
<td>4</td>
<td>20.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3-5 hours</td>
<td>9</td>
<td>45.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5-7 hours</td>
<td>6</td>
<td>30.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>over 7 hours</td>
<td>1</td>
<td>5.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of online courses previously participated in</th>
<th>Time of using Internet (h/day)</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0-1 years</td>
<td>1</td>
<td>5.00%</td>
</tr>
<tr>
<td>1</td>
<td>1-3 years</td>
<td>1</td>
<td>5.00%</td>
</tr>
<tr>
<td>2</td>
<td>3-5 years</td>
<td>3</td>
<td>15.00%</td>
</tr>
<tr>
<td>5</td>
<td>5-7 years</td>
<td>10</td>
<td>50.00%</td>
</tr>
<tr>
<td>4</td>
<td>over 7 years</td>
<td>5</td>
<td>25.00%</td>
</tr>
</tbody>
</table>

The evaluation followed a pre-test and post-test implemented in the following steps. First, instrumentnation of the evaluation tool was established. The items measuring the five constructs were adopted from related literature (Arbaugh, 2000; Davis, 1989; Venkatesh & Davis, 2000). Responses to the items were ranked on a 5-point Likert scale (1 represented “strongly disagree” and 5 represented “strongly agree”). Second, before the students started to use the system, a pre-test questionnaire was administered to gauge their attitudes and intentions regarding using online learning systems. The pre-test questionnaire also collected the subjects’ background information, e.g., gender, age, and Internet and online learning experiences (see Table 1). Third, after using the learning system, a post-test questionnaire was administered to collect students’ feedback. This measured students’ perceptions concerning ease of use of, usefulness of, and satisfaction with the system, as well as their post-use attitudes and intentions regarding using online learning. Four, quantitatively exploratory data analyses were implemented to obtain the results. Finally,
interviews were conducted to collect qualitative feedback from the participants regarding their comments on the system and perceived effectiveness of the system on learning.

**Survey Results**

Students’ responses to the survey were analyzed in the following steps. *First*, Paired-Samples T-tests of pre-test and post-test attitudes and intention were implemented to check the overall effectiveness of the system in changing students’ attitudes and intention regarding using online learning systems. The assumption was that students’ post-test attitudes and intentions would improve compared to those of the pre-test if the current system was effective overall. *Second*, Boxplots were reported describing the exploratory characteristics, including minimum, maximum, median, mean, and the first and third quantiles of perceived ease of use and usefulness of the overall system and the seven functions. *Third*, the correlations between perceived ease of use and usefulness of the overall system and perceived ease of use and usefulness of the seven functions were presented in bar charts. *Four*, Boxplots depicting the characteristics of students’ satisfaction with the overall system, the seven functions, and other three core elements of online courses were reported. In addition to system-related functions, other core elements of online courses such as learning contents, learning support (from tutors or peer learners), and technical support were highly integrated with the online learning environment, influencing learners’ satisfaction (Ozkan & Koseler, 2009; Iivari, 2005). These elements were therefore included in the data analysis. *Finally*, the correlations between students’ satisfaction with the overall system and students’ satisfaction with each individual function and element were given. The results of the statistical analyses are presented as follows.

**Attitudes and intention to use**

Table 2 reports the results of Paired-Sample T-test for students’ pre-test and post-test attitudes towards using online learning systems. The results suggest that using the JAVA e-Teacher system had a positive impact on students’ attitudes towards online learning, which signifies that the learning experiences with the current system were clearly positive.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT-post</td>
<td>4.32</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT-pre</td>
<td>3.82</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AT-post) – (AT-pre)</td>
<td>0.50</td>
<td>0.72</td>
<td>3.10</td>
<td>19</td>
<td>0.006</td>
</tr>
</tbody>
</table>

a: post-test attitude; b: pre-test attitude

Table 3 shows the results of Paired-Sample T-test for students’ pre-test and post-test intention to use online learning systems. The results show that after using the system the students tended to be more willing to use online learning systems, which means that the current system was effective in enhancing students’ online learning intention.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITU-post</td>
<td>4.72</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITU-pre</td>
<td>3.95</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ITU-post) – (ITU-pre)</td>
<td>0.77</td>
<td>0.85</td>
<td>4.02</td>
<td>19</td>
<td>0.001</td>
</tr>
</tbody>
</table>

a: post-test intention to use; b: pre-test intention to use

**Perceived ease of use**

Figure 6 presents the exploratory characteristics of perceived ease of use of the overall system and the seven functions. The Boxplot shows that the overall system and the functions were perceived to be easy to use
The function for uploading Q&A messages with links to knowledge concepts was perceived as the easiest to use, and the knowledge map as the second easiest to use.

Figure 6. Boxplot of exploratory characteristics of perceived ease of use.

Figure 7. Correlations on perceived ease of use

Figure 7 describes the Pearson correlations between perceived ease of use of the overall system and perceived ease of use of the functions. A high correlation denotes a high degree of co-variation and association between perceived ease of use of the overall system and perceived ease of use of specific functions. Figure 7 suggests that perceived ease of use of the functions for uploading Q&A messages with links to knowledge concepts (Q&A1-EOU), self-tests around knowledge concepts (TEST-EOU), and guidance on learning navigation in maps (GUI-EOU) are most highly related to perceived ease of use of the overall system.
Perceived usefulness

Figure 8 presents the exploratory characteristics of perceived usefulness of the overall system and its functions. The Boxplot shows that all the seven functions were perceived to be useful for online learning (means>4), and the functions of knowledge maps, self-tests around knowledge concepts, and guidance on learning navigation in maps, were reported to be very useful.

Figure 8. Boxplot of exploratory characteristics of perceived usefulness

<table>
<thead>
<tr>
<th>Function</th>
<th>Q1</th>
<th>min</th>
<th>median</th>
<th>mean</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge map</td>
<td>4.38</td>
<td>3.00</td>
<td>5.00</td>
<td>4.55</td>
<td>5.00</td>
</tr>
<tr>
<td>Self-test</td>
<td>4.44</td>
<td>3.00</td>
<td>5.00</td>
<td>4.64</td>
<td>5.00</td>
</tr>
<tr>
<td>Guidance</td>
<td>4.00</td>
<td>2.00</td>
<td>4.00</td>
<td>4.53</td>
<td>5.00</td>
</tr>
<tr>
<td>Chat&amp;consultation</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.20</td>
<td>5.00</td>
</tr>
<tr>
<td>Q&amp;A:upload</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.20</td>
<td>5.00</td>
</tr>
<tr>
<td>Q&amp;A:presentation</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.15</td>
<td>5.00</td>
</tr>
<tr>
<td>Discussion</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.08</td>
<td>5.00</td>
</tr>
<tr>
<td>Overall</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.63</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Figure 9 reports the Pearson correlations between perceived usefulness of the overall system and perceived usefulness of the functions. The bar chart shows that perceived usefulness of the functions for self-tests around knowledge concepts (TEST-U), knowledge maps (MAP-U), and guidance on learning navigation in maps (GUI-U) were highly related to perceived usefulness of the overall system.

Figure 9. Correlations on perceived usefulness
Satisfaction

Figure 10 presents the exploratory characteristics of students’ satisfaction with the overall system, with the seven system-related functions, and with other three core elements of online courses (course content, learning support from tutors and peers, and technical support). The Boxplot shows that the students’ responses to satisfaction with all the aspects were very positive. In particular, the students were highly satisfied with the function of guidance on learning navigation in maps and learning support provided by tutors and peers in the online course.

![Boxplot of exploratory characteristics of satisfaction](image)

**Figure 10.** Boxplot of exploratory characteristics of satisfaction

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>min</th>
<th>median</th>
<th>mean</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>knowledge map</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>self-test</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>guidance</td>
<td>4.50</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>chat&amp;consult</td>
<td>4.38</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Q&amp;A:upload</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Q&amp;A:presentation</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>discussion</td>
<td>4.25</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>content</td>
<td>5.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>learning support</td>
<td>5.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>technical support</td>
<td>5.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
<tr>
<td>overall</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

**Figure 11.** Correlations on satisfactions

![Correlations](image)
Figure 11 presents the Pearson correlations between students’ satisfaction with the overall system and students’ satisfaction with the system-related functions and course-related elements. The chart shows that students’ satisfaction with the functions for self-tests around knowledge concepts (TEST-U), guidance on learning navigation in maps (GUI-U), and online chat and consultation with tutors for conceptual and content understanding (CHAT-SA) were highly associated with overall satisfaction. In terms of course-related elements, students’ satisfaction with course content and learning support were found to be strongly related to their satisfaction with the system.

Findings from the Interview

Nineteen out of the twenty participants attended the interview. Each participant was interviewed individually and anonymously for their comments on the system and perceived impact of the system on their learning. Students were encouraged to discuss any issue they felt was relevant. The findings from the interview are reported as follows.

All the participants expressed their clear satisfaction with the system. As one said, “JAVA e-Teacher is the best learning system that I have ever used.” In particular, they expressed their strong preference for visualized knowledge maps. They felt that key knowledge concepts of the course were clearly outlined in the maps, providing learners with a preliminary understanding of the course before they started to learn. One student said: “The maps are like a reference system, making it easy for us to engage in a new learning environment with more flexibility.” Students mentioned that presenting knowledge components and their links in visual maps helped them memorize important concepts and facilitate high order thinking. As one remarked, “What we learnt from several maps is not less than what we learnt from a large number of texts.” Another said, “Learning is usually boring in many situations; however, visual knowledge maps with well-designed and user-friendly interfaces make our learning enjoyable and more fun.” While many students liked to drag and move knowledge concepts in a map, one participant felt that this might mess up the presentation by improper moves.

In relation to other functions, most participants commented that it was very convenient to browse the knowledge maps and access learning resources by clicking any topic they want to learn. As one said, “In case I want to know more about specific knowledge, I just click it. It is so simple and direct, saving both time and effort.” Participants also recommended that learning guidance generated by the system was helpful. One said, “The system tells me what is basic knowledge, or what I should learn at the current stage. This helps novices like me feel comfortable and learn easily.” They also enjoyed the color change in knowledge maps which reflected their individual learning status. Almost all the students liked to use the self-tests and discussion forum. One said: “The tests can pinpoint my weak points at anytime, instead of at the end of the course. This helps me consolidate my knowledge step by step.” Another said: “I enjoy visiting the Q&A forum, especially using my knowledge to answer other students’ questions, which finally improves my own learning.”

At the same time, the participants mentioned some limitations of the system. They expected more practical examples, simulations, different levels of learning materials, and NCRE exam papers to be included in the system. They also suggested using other types of assessment in addition to multiple-choice questions.

Discussion and Conclusion

This study proposed a KV approach to the problem of cognitive overload and conceptual and navigational disorientation in a resource-abundant and self-regulated online learning environment. The approach was investigated through a case study of an online course. The investigation involved the design, development, and evaluation of an enhanced learning system “JAVA e-Teacher” by using the proposed approach. The focus was on visualization of domain knowledge structure and integrating the structure with curriculum design, learning resources, learning assessment, intellectual process, and social learning. Students reported a high degree of satisfaction with and perceived ease of use and usefulness of the system and its functions relevant to KV. Moreover, students’ attitudes and intentions regarding using online learning systems were found to have improved after they used the JAVA e-Teacher system, which indicates the overall effectiveness of the system. At the same time, students reported positive feedback in the interviews regarding the effect of the system on learning such as scaffolding conceptual understanding, improving memorization and thinking, facilitating access to learning resources, and supporting individual and social learning.
The evaluation results reflect the success of the system in terms of user satisfaction and acceptance, and effectiveness for learning. In addition, the correlation analyses found that “knowledge maps”, “Q&A discussion around knowledge concepts”, and “online chat for conceptual and content understanding” were important factors in perceived usefulness, ease of use, and satisfaction concerning the overall system respectively; while “learning guidance for navigation” and “self-tests around knowledge concepts” were found important for all the three evaluation constructs. While all these KV related functions played critical roles in the success of the overall system, each of them relied on the construction of visualized knowledge structure, the key element of the proposed approach. On the other hand, learning support for learners and learning contents or information quality (Alkhattabi et al., 2010) were found to be strongly correlated to user satisfaction in this study, a finding to which online learning designers and developers should pay high attention.

Although the study was conducted with a programming course, the approach was directly applicable to other learning programs where the domain knowledge structures can be specified. The evaluation results reported in this paper are limited due to the small sample size. Findings will be revisited in a number of continuous experiments with more participants using the system. Further investigation on the effectiveness of the proposed approach, in particular its effect on reducing cognitive load and enhancing self-regulated learning will be reported in further studies.

Acknowledgements

This research is supported by a UGC GRF Grant (No. 717708) from the Hong Kong SAR Government, a Seeding Fund for Basic Research (No. 200911159142), a Seed Fund for Applied Research (No.201002160030), and a Earmarked Teaching Development Fund from The University of Hong Kong. The authors thank Professor Haijing Jiang for his valuable comments to this project.

References


Numerous researchers have experimented different ways of teaching and found serious problems in geometry their memory (Sweller, 1988). This often causes cognitive overload and poses a negative effect on students’ learning.

When a teacher discusses a geometry proof problem in class, it generally involves oral presentation of a formal proof and body movements pointing at different parts of the figure of the problem. Students must watch, listen, jot notes, and think as a lecture proceeds. They have to refer to many elements of the instruction and incorporate them into memory.

One characteristic of human intelligence is the use of different types of representation. NCTM (1989, 2000) has advocated a K-12 curriculum that stresses mathematical connections among multiple representations. Janvier (1987a) used an interesting analogy to explain the application of multiple representations in teaching: when using multiple representations to illustrate a mathematical concept, it is like a star-like iceberg. The sharp angle of each star indicates a representation, and the students’ learning objective is to construct an entire star-like iceberg. Teaching, meanwhile, is to let the star-like iceberg emerge from the water so that students can flexibly operate each representation.

Behr et al. (1983) built on Bruner’s (1966) representation theory to propose an interactive model using a representational system to improve the learning of mathematics. Problem solving would be an easier task if a student is aware of how the structures of different representations are interactively connected. The teaching of multiple representations should be emphasized. As for mathematics education, Janvier (1987b) proposed that instruction on functions could make use of verbal description, table, formula, and graph representations. Kieran (1993) proposed similar representations of functions: situation (graphical or textual description), table, diagram, and algebraic equations. QUADRATIC (Wood & Wood, 1999), using the area of squares to make salient the geometric properties of algebraic expressions, was designed to teach students about equivalences implicit in mappings between algebraic and geometric representations. Wong et al. (2007a) studied the use of representations of natural language, telegraphic
description and diagram in problems about area and circumference of geometric shapes for students of elementary school.

In the mid-1980s to the 1990s, some computer programs are used in teaching and exploring geometry in school. In this study, we divided those computer systems into two categories, one focused on Dynamic Geometry Environment (DGE), which used for exploration; the other is Expert System (ES) or Intelligent Tutoring System (ITS), which is used for problem-solving tasks. Some popular DGEs include Geometer's Sketchpad (GSP; Jackiw, 1997), Cabri Geometry II (http://www.cabri.com/v2/pages/en/index.php), Geometry Expert (Chou et al., 1996), and Cinderella's Café (http://www.cinderella.de/tiki-index.php). While these DGEs share one common focus on dynamic geometry figures, Geometry Expert also does automatic theorem proving. In a dynamic geometry figure, students can drag a geometry object such as a vertex of a triangle and change the figure dynamically while preserving the given conditions of the figure and geometric invariants, which are the consequences of the given conditions. Thus these programs are commonly used to demonstrate geometry theorems.

A well-known ITS Geometry Proof Tutor (GPT) was developed at Carnegie Mellon University to support students in writing Euclidean proofs, and results showed that the tutors were roughly half as effective as a human tutor (Epstein & Hillegeist, 1990). A similar problem-solving environment is ANGLE, where students can construct graphical representations of Euclidean proofs. Unlike GPT, ANGLE is based on a cognitive model that models and therefore facilitates thinking with an expert-like representation of target knowledge (Koedinger & Anderson, 1993b). While GPT and ANGLE help students produce a complete proof, a web-based system is designed to enhance students’ reading and learning of proofs in this study.

Multiple Representations

Some scholars (Lesh, Post & Behr, 1987; Ainsworth, 1999; Duval, 1998) noted that students can learn to do translation between multiple representations and make transitions within each representation. As a result, students can modify incorrect concepts and conjectures by comparing multiple perspectives provided by different representations. However, some researchers have found that even though teachers alert students to the importance of understanding each type of representation, students can fail to notice regularities and discrepancies between representations (Kaput, 1987; de Jong et al., 1998; Kozma, 2003). Sweller (1988) pointed out that when learning with several representations, learners are required to relate disparate sources of information, which may cause a heavy cognitive load that leaves few resources for actual learning. Seufert (2003) noted that learners with little prior knowledge are unable to acquire knowledge from multiple representations. She recommended the provision of more semantic aids in order to reduce such cognitive load.

In this study, we adopt the DeFT (Design, Functions, Tasks) framework of learning with multiple representations (Ainsworth, 1999, 2006; Ainsworth & Van Labeke, 2004; Van Labeke & Ainsworth, 2001). The DeFT framework refers to different pedagogical functions that multiple representations can play, the design parameters unique to learning with multiple representations and the cognitive tasks undertaken by a learner interacting with multiple representations. There are three main functions of multiple representations: complement, constrain and construct. Multiple representations complement each other by supporting complementary processes or conveying complementary information. Two representations constrain each other, as one supports interpretation of the other. Finally, multiple representations can be used to encourage learners to construct a deeper understanding of a situation.

When learning with multiple representations, learners have to deal with complex cognitive tasks Van Labeke & Ainsworth (2001) and van der Meij & de Jong (2006). First of all, learners have to understand the syntax of each representation. Second, learners must also understand which parts of the domain are represented. Third, learners have to relate the representations to each other whether the representations present the same information. Fourth, learners have to translate between the representations. As Ainsworth (2006, p. 195-196) argued, “DeFT clarifies the pedagogical functions that multiple representations serve, the often-complex learning demands that are associated with their use and in so doing aims to consider the ways that different designs of multi-representational systems impact upon the process of learning.”

Although there are many empirical studies of computer-based learning environment based on multiple representation theory (e.g. Wood & Wood, 1999; Van Labeke & Ainsworth, 2001; van der Meij & de Jong, 2006; Brenner, et al.,
1997; Hsu, 2006), geometry proof learning is seldom discussed from the perspective of such theory. In this study, we use the DeFT framework to develop a computer-based geometry proof learning environment and investigate its effects on learners.

**Representations for Geometry Problems and Proofs**

Healy and Hoyles (1998) and Duval (2002) emphasized that in order to construct a geometry proof, students should first identify the given information and then search for critical hints. They should then apply appropriate geometric properties to conduct deductive reasoning, organize all deductive steps, and write down the sequence of steps inferring from the given conditions to the conclusion. Various representations can be used to specify a geometry problem, support the reasoning process, and represent the final proof.

A geometry problem is specified with a verbal description, often accompanied by a figure. As Mayer and Sims (1994) pointed out, students can build more referential connections when verbal and visual materials are presented contiguously than when they are presented separately. Clements and Battista (1992) also noted that for a student to successfully prove a theorem, the student must build semantic links between the concepts of geometry and the features of a figure. Through bi-directional connections, students can clearly demonstrate the interrelation between the geometric components in a verbal description and the objects in an accompanied figure. Schnotz’s (2002) integrative model of text (descriptive representation) and picture (depictive representation) comprehension emphasizes that good graphic design is crucial for individuals with low prior knowledge who need pictorial support in constructing mental models.

Duval (1998) proposed that learning geometry involves three kinds of cognitive processes: visualization, construction, and reasoning. The foundation of reasoning lies in mathematical language and logic. Although the visualization process can sometimes provide individuals with heuristics in proving a theorem, a figure can be misleading sometimes because students are often influenced by paradigmatic images. Therefore, simple-minded visualization can interfere with the process of deduction. For example, students may take a rotated square as a “non-square rhombus.” Duval argued, however, “these three kinds of cognitive process are closely connected and their synergy is cognitively necessary for proficiency in geometry” (p. 38). In other words, reasoning and visualization processes can complement each other. But, in order to avoid the influence of the paradigmatic images, we can make use of a dynamic geometry figure. Duval (2000) suggested that only through the coordination of different representations could students develop geometry problem solving skills. Previous studies indicate that three types of representations should be useful for learning geometry proofs. The following presents a detailed explanation of these representations.

**Problem Representation**

The first type of representation is the problem itself, or the problem representation (Fig. 1), which is generally expressed as a text. The problem representation specifies some given conditions and a goal that needs to be proved. Students need to understand the mathematical symbols and language, and the logical relationship between the given conditions and the goal condition.

**Visual Representation**

The second type of representation is visual representation (Fig. 1), which can be a static or dynamic figure. A static figure groups together relevant information from a problem representation. The correspondence between a text and its figure can provide complementary information for students, which could enhance their comprehension of the geometry problem. A proof involves inferences from the given conditions to a conclusion with a sequence of deductions and a simple figure provides little help for students to construct a proof. However, students can visually check whether a proposition is valid in a figure. In other words, static figures can serve the role of constraining an interpretation.
A dynamic geometry figure is also a visual representation. A student can interact with a dynamic figure, which provides a clear picture of abstract mathematical ideas through concrete object dragging. By manipulating a dynamic figure and observing how it changes, students may be able to avoid over-generalization of theorems from paradigmatic images. DGE allows students to either falsify propositions or enhance the degree to which propositions are believable. The invariant nature of the geometric properties of a dynamic figure can constrain the interpretation of the propositions in a formal proof.

**Proof Representation**

The third type of problem representation is that of a proof (Fig. 1), which can be a formal proof or a proof tree. As the most rigorous form of solution in geometry proof, a formal proof is composed of a sequence of deductive steps. Formal proof is also the standard deductive reasoning format that instructors and students use to communicate with each other. As Yang and Lin (2008) pointed out, when reading a proof, a learner needs to recognize the role of an example figure, identify the given conditions and the goal condition, generate implicit hypothesis or properties, and apply inference rules. Indeed, the comprehension of a proof is a complicated cognitive process. In order to reduce the cognitive load of proof comprehension on learners, we propose the use of a proof tree.

A proof tree (Fig. 2) is another representation of a proof. ANGLE (Koedinger & Anderson, 1993a) and Matsuda & VanLehn (2005) utilize a tree-like structure to represent a network of inferences. As a visual representation of a formal proof, a proof tree is a hierarchy of nodes inferring from some given conditions to the goal condition (Wong, et al., 2007b). The root node of a proof tree is the goal condition to be proved. The tree consists of leaf nodes and derived nodes. Each node of a proof tree represents a proof step. Each leaf node is a given fact or self-evident condition, e.g., a segment’s length equals that of itself. A proof tree shows the logical structure underlying a common textbook proof. In a proof tree, the logical relation between the given facts and the goal is made explicit graphically.

In Fig. 2, the top is the description of a problem, the lower left is a formal proof, and the lower right is a proof tree. In a top-down approach, the proof tree presents the proof process and highlights the logic relationship between the given conditions and the goal. By studying a proof tree, students can check the entire structure of the inferences leading to the goal.

Addressing the difficulties that students encounter in learning geometry, this study proposes a multimedia learning environment to let students interact with multiple representations relevant to a geometry proof. This interactive learning environment is called MR Geo (Multiple Representations for Geometry).
MR Geo

Consider the graphical user interface of MR Geo. There are four frames with five representations relevant to geometry proof (Fig. 3). These representations include problem description, static figure, dynamic geometry figure, formal proof, and proof tree.

**Figure 2. A formal proof and its proof tree**

1. ABCD is a parallelogram  (Given)
2. ∴ ABCD is a parallelogram, ∴ \( \overline{AB} \parallel \overline{DC} \)  (Def. of parallelogram)
3. ∴ ABCD is a parallelogram, ∴ \( \overline{AD} \parallel \overline{BC} \)  (Def. of parallelogram)
4. \( \overline{AB} /\!/\overline{DC} \) , ∴ \( \angle ACD = \angle CAB \) (Alt. int. angles)
5. ∴ \( \overline{AD} /\!/\overline{BC} \) , ∴ \( \angle DAC = \angle BCA \) (Alt. int. angles)
6. \( \overline{AC} = \overline{AC} \)  (Reflexive law)
7. \( \triangle ABC \cong \triangle CDA \)  (ASA)

**Figure 3. A screenshot of MR Geo**

Problem description (A) and static figure (B): In Fig. 3, the top left frame of the interface shows the verbal description a geometry problem accompanied by a static figure. When a user clicks on the problem text, formal proof or proof tree, the clicked objects or the objects in a proof step will be highlighted in the figure with different colors. This design can enhance users’ understanding of the correspondence between the figure and other representations.
Dynamic geometry frame (C): The top right frame shows an embedded GSP’s Java applet. Students can drag geometric objects in the figure with the mouse. This will cause the values of some chosen attributes of this figure to change dynamically and the values will be displayed at the same time. In this way, students can observe the invariant properties of the dynamic geometry figure and test the validity of each proof step visually.

Formal proof frame (D): The bottom left frame shows a formal proof. A proof is presented in a two-column format. Each row specifies the conditions, which can be given or propositions derived earlier, for deriving a proposition with an inference rule.

Proof tree frame (E): The bottom right frame shows a proof tree. With a click on any node on the proof tree, a student can see the corresponding inference step highlighted in the formal proof. This will help students understand the correspondence between the linear, formal proof and the proof tree. The proof tree clearly shows the relationship between each parent and its children nodes.

In MR Geo, when an object (such as verbal, proposition, and tree’s node) is selected in one representation, the corresponding objects in the other representations will be highlighted. For example, consider the problem “Given a parallelogram ABCD with diagonal \( \overline{AB} \). To prove: \( \triangle ABC \cong \triangle ADC \).” When students choose the problem’s given condition “Parallelogram ABCD” in the problem representation, the line segments \( \overline{AB}, \overline{BC}, \overline{CD}, \) and \( \overline{AD} \) will be highlighted with the same color in the static figure. In the formal proof, the first row of “Parallelogram ABCD” would also be marked, as well as the two “Parallelogram ABCD” nodes in the proof tree. MR Geo provides simultaneous flashing and color-coding to support effective dissemination of information. If students doubt the validity of \( \angle ACD = \angle BAC \), she can drag any vertex of the parallelogram on the dynamic representation to check whether the condition is always true.

**Purpose of the Study**

Given the success of using multiple representations for math instruction in past studies, this study investigates how students would react to the use of multiple representations such as proof tree, formal proof and dynamic geometry figure when they read a proof and construct a proof in the system. In particular, there are several questions we would like to answer. Which representations do students prefer when they perform tasks related to theorem proving in MR Geo? After using MR Geo, will students’ attitudes towards learning geometry undergo any change? Which types of students benefit more from multiple representations in MR Geo?

**Methodology**

In order to understand the effect of MR Geo on junior high school students’ learning of geometry proofs, we conducted an experiment that spanned six weeks. This experiment used materials on parallelograms and triangles from textbooks and reference books for grade nine mathematics that covered basic algebraic concepts (such as equality axiom), properties and elements of geometry (such as perpendicular bisectors, angle bisectors), and similarity and congruence of triangles.

**Participants**

The participants were 96 ninth grade students selected from three classes in two public junior high schools in southern Taiwan. There were forty-seven males and forty-nine females with ages ranging from 14.5 to 16.0 (average 15.01). They had already learned basic geometric concepts and elementary proofs. All participants possessed the basic skills to work with a computer. Moreover, all participants did not use any computer-assisted learning system in the learning of geometry and its proof.

As a pretest, we adopted Standard Progressive Matrices Plus (or SPM+; Raven, 2004) to assess students’ figure deduction skill. The SPM+, which is a standardized test for students in Taiwan, keeps the 60-item, format of the
Classic SPM. Its test-retest reliability was .87 and split-half reliability was .80-.86. Also, its correlation with math grades was .65. Overall speaking, the reliability and validity of SPM+ are satisfactory.

Based on their mean scores of SPM+ and math grades with equal weighting, the students were divided into three groups: high-achievement group (or HG; 28 persons), medium-achievement group (or MG; 40 persons), and low-achievement group (or LG; 28 persons). By splitting the students into three groups, we can find out which types of learning problems would be encountered by each group of students.

**Materials**

Research tools used in this study included MR Geo and a questionnaire survey. We designed four tasks of practice and testing for learners in MR Geo, progressing from simple tasks to complex tasks. In simple tasks, most information was provided for the learners while in complex tasks, little information was provided and more efforts were expected from the learners. The four tasks are shown in Fig. 4.

![Screenshot of four tasks by students](image)

(a) Finding missing nodes  
(b) Local theorem proving  
(c) Translation of a proof tree into a formal proof  
(d) Concept testing

*Figure 4. Screenshots of four tasks by students*

Task 1 – Finding missing inference steps: To recover the missing nodes of a proof tree, the students can interact with the problem, the formal proof, the static figure, or the dynamic geometry figure (see Fig. 4(a)). By observing the mutual correspondence among these representations, the students might improve their comprehension of the formal proof. In particular, if a student can translate the formal proof into a proof tree, she should be able to recover the missing nodes.
Task 2 – Local theorem proving: Some steps in the formal proof are missing. In the form of multiple-choice questions (select one of 4 choices), the system asks students to fill in the missing deductive steps. In Fig. 4(b), students can obtain relevant information by reading the proof tree and dragging the dynamic figure to check candidate propositions. The main purpose of this task is to help students complete a formal proof after they have read and understood the proof tree.

Task 3 – Translation of a proof tree into a formal proof: Students are asked to construct a formal proof by dragging nodes of the proof tree and drop into rows of the formal proof frame (see Fig. 4(c)). Moreover, they can reorder the rows by dragging them around. In this way, they can explore different orderings of the steps. They can interact with the problem representation or the visual representations whenever they want to check any geometric features or propositions. The major purpose of this task is to improve students’ understanding of how to write a complete formal proof.

Task 4 – Concept testing: The main purpose of this task is to test students’ basic geometric concepts. Students are asked to pick geometric objects from a figure or answer multiple-choice questions. Consider the question in Fig. 4(d) as an example: “Please indicate on the figure a pair of alternate interior angles in the top right frame of the screen.” Students can use the mouse to select a geometric object such as “line”, “angle” by clicking at the object on the static figure. In Fig. 4(d), \( \angle DAC \) and \( \angle BCA \) are marked by a student as a pair of alternate interior angles.

Results and Discussion

Students’ Interactions with Representations in MR Geo

When students took a test with MR Geo in the last week of the experiment, they could interact with any of the five representations by pointing the mouse at a geometric object or a proof step. Table 1 shows how many times students interacted with the representations. An interaction was counted if the mouse stayed inside the frame of representation for over 2.5 seconds. The data showed that students interacted with proof tree most frequently (2.73 times), followed by formal proof (2.61 times).

<table>
<thead>
<tr>
<th>Representation</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem</td>
<td>1.51</td>
</tr>
<tr>
<td>Dynamic geometry figure</td>
<td>.77</td>
</tr>
<tr>
<td>Formal proof</td>
<td>2.61</td>
</tr>
<tr>
<td>Proof tree</td>
<td>2.73</td>
</tr>
</tbody>
</table>

Also, students interacted less frequently with problem representation (1.51 times) and dynamic geometry figure (.77 times). In problem representation, students usually spent little time to review the question after they have read and understood the given conditions and goal condition. Although students could consult the static figure to check given conditions, they were unable to manipulate the static figure freely like dynamic geometry figure. The interaction frequency of dynamic geometry figure was significantly lower than the other three representations.

Table 2. ANOVA test of students’ frequency of interaction with representations

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
<th>F</th>
<th>Scheffé test</th>
</tr>
</thead>
<tbody>
<tr>
<td>HG</td>
<td>28</td>
<td>3089</td>
<td>110.32</td>
<td>149.86</td>
<td>7.79*</td>
<td>MG&gt;HG</td>
</tr>
<tr>
<td>MG</td>
<td>40</td>
<td>4795</td>
<td>119.88</td>
<td>134.68</td>
<td></td>
<td>MG&gt;LG</td>
</tr>
<tr>
<td>LG</td>
<td>28</td>
<td>3111</td>
<td>111.11</td>
<td>90.62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: p< .05

Table 2 shows the three groups students’ interactions with representations in MR Geo in the test. In answering six proof questions, HG students interacted with representations for an average of 110.32 times, MG students 119.88 times, and LG students 111.11 times. These numbers showed that MG students interacted with representations most frequently, followed by LG students and HG students. ANOVA analysis demonstrated that F value was 7.79 (p<.00075). Scheffé post-hoc test was then used to compare the differences among groups, showing that the
interaction frequency of MG students was significantly higher than those of HG and LG students. Also, the difference between HG and LG groups was insignificant. Since MG students interacted more with multiple representations, they had more chances to build up the connections between these representations. Their learning may benefit most from these interactions. This is an interesting problem worthy of further study.

**Students’ Preferences of Representations**

After performing one task each week for four weeks, students became more familiar with MR Geo and the features of each representation. In the fifth week, the authors conducted a survey on students’ attitudes towards geometry learning in general and MR Geo in particular. The students were asked “What type(s) of representation(s) do you think help your understanding of geometry proof?” In answering, they could choose one or more of the five representations. According to the overall results in Table 3, many students preferred proof tree, formal proof, and problem description.

**Table 3. Preferences on representations in different groups of students**

<table>
<thead>
<tr>
<th>Representation</th>
<th>Overall No. of HG students</th>
<th>No. of MG students</th>
<th>No. of LG students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem</td>
<td>49 (51.0%)</td>
<td>8 (28.6%)</td>
<td>21 (52.5%)</td>
</tr>
<tr>
<td>Static figure</td>
<td>42 (43.8%)</td>
<td>9 (32.1%)</td>
<td>22 (55.0%)</td>
</tr>
<tr>
<td>Dynamic geometry figure</td>
<td>47 (49.0%)</td>
<td>10 (35.7%)</td>
<td>20 (50.0%)</td>
</tr>
<tr>
<td>Formal proof</td>
<td>61 (63.5%)</td>
<td>19 (67.9%)</td>
<td>31 (77.5%)</td>
</tr>
<tr>
<td>Proof tree</td>
<td>64 (66.7%)</td>
<td>18 (64.3%)</td>
<td>31 (77.5%)</td>
</tr>
<tr>
<td>None of the above</td>
<td>4 (4.2%)</td>
<td>0</td>
<td>1 (2.5%)</td>
</tr>
</tbody>
</table>

By focusing on formal proof and proof tree in Table 3, we found that HG and MG students liked the formal proof and proof tree more or less the same. This indicates that HG and MG students might know that these two representations express similar information and they can be translated into one another. However, for LG students, a graphical proof tree seems to be more revealing and a better alternative to a formal proof.

Students’ relative preferences in dynamic geometry figure are quite interesting. For both HG and MG students, dynamic geometry figure is much less attractive than formal proof and proof tree. In contrast, LG students preferred dynamic geometry figure, formal proof and proof tree in this order. Our preliminary interpretation of the results from Tables 1 and 3 is as follows. HG students could understand formal proof and proof tree well enough that they did not need to consult the dynamic figure. MG and LG students did not know how to check the validity of geometric conditions with a dynamic figure. LG students liked to play with a dynamic figure because they were attracted by the intriguing dynamics. Although playing with a dynamic figure did not help their understanding of a proof in general, a DGE could make geometry more interesting and attractive to them.

**Students’ Reactions to MR Geo**

After using MR Geo, the students filled out a questionnaire to express their opinions of MR Geo. Independent t tests were made between the mean score of any pair of three groups for all items. Analysis indicated that there was significant difference in five items. The first item was “Did you enjoy using MR Geo?” The mean scores of MG and LG students was significantly higher than that of HG (t = 2.74, 3.68 for MG-HG and LG-HG, p<.01 and p<.001 respectively). The second item was “In MR Geo, DGE can be used to measure the properties of geometry figures. It can also help you understand or validate the steps of the proof.” The mean score of HG students was significantly higher than those of MG and LG (t = 3.27, 5.67 for HG-MG and HG-LG respectively, all p<.001). The third item was “In MR Geo, DGE can be used to measure the properties of geometry figures. It can also help you understand or validate the steps of the proof.” The mean score of HG students was significantly higher than those of MG and LG (t = 3.27, 5.67 for HG-MG and HG-LG respectively, all p<.001). The fourth item was “By observing the proof tree structure, you can identify whether proof steps can be swapped.” The mean scores of HG and MG students were significantly higher than that of LG (t = 5.42, 4.45 for HG-LG and MG-LG respectively, all p<.001). The fourth item was “By observing the proof tree structure, you can identify whether proof steps can be swapped.” The mean scores of HG and MG students were significantly higher than that of LG (t = 5.25, 4.82 for HG-LG and MG-LG respectively, all p<.001). The fourth item was “Through the exercises of MR Geo, geometry proofs became more interesting to you.” The mean scores of MG and LG students were significantly higher than that of HG (t = 3.27, 3.64 for MG-HG and LG-HG respectively, all p<.001). The results also indicated that over 85% of the 96
ninth grade students agreed that proof tree can help them better comprehend a geometry proof. Table 4 shows the five items and their results.

Table 4. Students’ responses to questionnaire

<table>
<thead>
<tr>
<th>Statement</th>
<th>Options</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>No comment</th>
<th>Disagree</th>
<th>Strongly disagree</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did you enjoy using MR Geo?</td>
<td>19 (19.8%)</td>
<td>51 (53.1%)</td>
<td>25 (26.1%)</td>
<td>1 (1.0%)</td>
<td>0</td>
<td>3.92</td>
<td></td>
</tr>
<tr>
<td>2. In MR Geo, DGE can be used to measure the properties of geometry figures. It can also help you understand or validate the steps of the proof.</td>
<td>28 (29.2%)</td>
<td>44 (45.8%)</td>
<td>22 (22.9%)</td>
<td>2 (2.1%)</td>
<td>0</td>
<td>4.02</td>
<td></td>
</tr>
<tr>
<td>3. Overall, the MR Geo can help you read and comprehend geometry proof problems.</td>
<td>28 (29.2%)</td>
<td>56 (58.3%)</td>
<td>11 (11.5%)</td>
<td>1 (1.0%)</td>
<td>0</td>
<td>4.16</td>
<td></td>
</tr>
<tr>
<td>4. By observing the proof tree structure, you can identify whether proof steps can be swapped.</td>
<td>34 (35.4%)</td>
<td>48 (50.0%)</td>
<td>12 (12.5%)</td>
<td>2 (2.1%)</td>
<td>0</td>
<td>4.19</td>
<td></td>
</tr>
<tr>
<td>5. Through the exercises of MR Geo, geometry proofs became more interesting to you.</td>
<td>24 (25.0%)</td>
<td>49 (51.1%)</td>
<td>22 (22.9%)</td>
<td>1 (1.0%)</td>
<td>0</td>
<td>3.97</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the questionnaire items, we also asked students to write down their comments on MR Geo, some of which are listed as follows. “This computer program helps me understand geometry proof. In future exams, I would attempt to solve geometry proof problems” (LG). “It makes me feel that geometry proof is not that difficult” (LG). “Although I did not fully understand the geometry proofs, I believe that MR Geo is good” (LG). “I like to play with the dynamic figure, it is so funny” (LG). “I feel that geometry proof is very interesting” (MG & LG). “Compared to textbooks, this is an easier way to understand geometry proof” (MG & LG). “Previously, I didn’t quite understand geometry proofs; I have now gradually developed a better understanding” (MG). “I think reading questions on paper is easier than looking at the computer display because the display confuses me” (HG). “This learning method is tiring” (HG). “DGE can help me check and verify some propositions (statements, relationships) in the formal proof or proof tree” (HG).

Conclusion

**MG students interacted more frequently with representations than HG and LG students**

HG students were more comfortable in using a formal proof so their interactions with other representations were not as frequent as MG students. LG students, with a weaker background in math, might not improve their performance by much even with more interactions with the representations. In this way, MG students might benefit most from their interactions with the representations in MR Geo.

**Multiple representations improved students’ perspective of geometry proof**

Problem description, static figure, dynamic geometry figure, formal proof and proof tree had different effects on students. In particular, problem description and static figure could assist students’ understanding of the problem context and only a few HG students said that dynamic geometry could help them confirm or reject a proposition. The connection between formal proof and proof tree raised students’ comprehension of geometry proof. Some LG students indicated that after understanding the geometry proving process, they no longer hated geometry classes. The above results indicated that MR Geo might offer an attractive, alternative approach to geometry education with multiple representations in a computer-assisted learning environment, comparing to traditional classroom teaching.

**LG and MG Students could understand a formal proof better with the help of a proof tree**

Proof tree offers an outline of a complete geometry proof. Understanding the isomorphism between proof tree and formal proof can unleash the complementary and interpretation functions of both representations. By using a proof tree at the same time, students could better understand the formal proof representation. In short, students believed that using a proof tree in addition to a formal proof had a positive effect on their learning of geometry proof. This finding has some implications on the instruction of geometry proofs. A teacher in a traditional classroom can use a
proof tree to accompany a formal proof. This simple addition of a new representation might inspire LG and MG students to develop a better understanding of geometry proofs.

Encouraged by the above preliminary results, we will continue to analyze students’ data on proving theorems collected in the same experiment described above. What cognitive processes did they go through during problem solving with different representations? Did they really benefit from these representations and performed better as they claimed? Another paper will try to answer these intriguing questions, which might provide invaluable insights for geometry teachers to motivate more students to develop greater interest, logical reasoning and analytic skills in geometry.

Acknowledgement

This work was supported by the National Science Council, Taiwan (NSC 98-2511-S-224-005-MY3 and NSC 98-2511-S-224-004-MY2). We would like to thank three anonymous reviewers and the editor for their comments.

References


Learning as ‘Knowing’: Towards Retaining and Visualizing Use in Virtual Settings

Demosthenes Akoumianakis
Department of Applied Informatics & Multimedia, Technological Education Institution (TEI) of Crete, Greece // da@epp.teicrete.gr

ABSTRACT
The paper elaborates on the assumption that in modern organisations collaborative learning is an enacted capability that is more about ‘acting’ and co-engaging in shared practices. In such settings, virtual learning can be conceived as an emergent knowledge process with no pre-determined outcomes that occupies multiple online and offline constituencies. Then, retaining and visualizing the history of co-engagement in practice becomes a pre-requisite for effective e-learning environments. Using two case studies exploring the social production of radically different information-based products, namely networked music performances and vacation packages, the present work frames virtual learning in the traces online ensembles leave as their members co-engage (both online and offline) in socially constituted activities. The analysis leads to several findings, with most prominent the following: (a) ‘Knowing’ is accomplished through recurrent co-engagement in a designated practice, (b) Learning in online settings emerges through the intertwining of online and offline activities, and (c) Artefacts of practice need to embody critical design qualities, if they are to serve learning as ‘knowing’ in practice.

Keywords
Virtual learning, Distributed collective practice toolkits, Visualization, Online tells

Introduction
Learning organizations rely heavily on distributed information processing, which is characterized by the norms and local practices of the sub-unit, customized styles of work and situational influence on problem solving. Learning in these settings is frequently framed away from the universal, the general and the timeless and more in relation to the local the particular and the timely (Toulmin, 1990; Suchman, 2000). To grasp the basics of such learning, researchers have explored a variety of theoretical constructs. Wenger (2001) uses the notion of communities of practice to coin ‘a group of people who share an interest in a domain of human endeavour and engage in a process of collective learning which strengthens sense of community’ (p. 1). The basic idea rests on anthropological perspectives that examine how adults learn through the performance of social practices rather than focusing on environments intentionally designed to support learning. Arguably, this type of learning is outside an instructor-student-course context and more likely to be encountered in settings where the community addresses not only the technical acquisition of skills required by a specific practice, but also the informal and social aspects of creating and sharing knowledge. It is in these communities of practice that people learn the intricacies of their job, explore the meaning of their work, construct an image of the organization, and develop a sense of professional self. Then, learning becomes an enacted capability, mediated by technological artefacts and enabled or constrained by the intrinsic configurations of people, artefacts and social relations.

Visual aids are widely recognised as amplifiers of the learning capabilities of humans used to foster sense-making and process complex information. Card, Macinlay & Shneiderman (1999) introduced the term information visualization to qualify a research area concerned primarily with the interactive manifestation of complex information structures. At the time, the focus was on computational manipulation of large data sets to extract value and amplify the users’ cognition. Since then, a variety of techniques have been developed for framing ambiguous states, bringing order to complexity, making sense out of seemingly unrelated things and finding insights that are buried in data. With the emergence of networked environments, information visualization was expanded in two prominent directions. The first, coined as social visualization, seeks to provide informative accounts of the social contexts in which information is created (Gilbert & Karahalios, 2009). The second, referred to as knowledge visualization, focuses on the ‘use of visual representations to improve the creation and transfer of knowledge between people’ (Eppler & Burkhard, 2004).

This paper presents an effort to exploit the synergistic use of both types of visualization to trace aspects of virtual learning in the history of recurrent co-engagement and the socially constituted activities of an online ensemble. The focus is on learning that emerges in practice and in the course of collaborative co-engagement in the social
production of artefacts whose historical design record pre-dates and frequently determines their tangible and material substance. Earlier works have convincingly demonstrated how this type of learning can be facilitated through virtual prototypes that pre-exist certain types of consumer goods (von Hippel & Katz, 2002). The current paper extends this body of knowledge into information-based product lines such as online vacation services that pre-exist the packages in which they are embedded and collaborative music rehearsals that pre-date the live performances or recorded audio files experienced by users. In all these cases, the material and the virtual substance of the artefact become strongly intertwined in a social production process that is organized around ‘construction-negotiation-reconstruction’ cycles. It is argued that throughout these cycles, learning stands out as an emergent knowledge process that can be revealed by (a) retaining social data i.e., electronic traces people leave as they go about their daily routine and patterns of interaction fostering competence building during social production and (b) using social visualization and knowledge visualization to detect facets of collective learning.

The paper is structured as follows. First, an effort is made to review the concept of ‘knowing’ in practice and discuss the role of and challenges for visualization in fostering learning as ‘knowing’ in practice. The following section presents distributed collective practice toolkits, emphasizing how such tools can exploit visualization to capture aspects of emergent knowledge processes. Then, two cases studies are reviewed where visualizations are employed as shared practice vocabularies capable of capturing online and offline collaborative behaviour, thereby unfolding a variety of informal learning patterns during network music performance and vacation package assembly, respectively. In light of the two cases, the discussion section consolidates critical issues for supporting knowing in practice and elaborates on implications for e-learning designers. The paper is wrapped up with a summary and conclusion.

**Background and Related Work**

Scholarship on organizational learning tends to create a sharp distinction between knowledge and ‘knowing’, highlighting some of the challenges confronting the process of ‘knowing’ in practice. This section reviews relevant works on the relationship between learning and the process of ‘knowing’. Through this lens, it is made possible to frame e-learning in the ‘virtual tells’ of the socially constituted practice of the collaborators and assess the role of visualization in facilitating reflective action in boundary spanning collaborative settings.

**Learning as ‘Knowing’ in Social Production**

The concept of knowing is evident in the works of several authors, including Schon (1983), Maturana and Varela (1998) and Orlikowski (2002). Schon in his field study of five professions argued that the skilful practice exhibited by professionals did not consist of applying some a priori knowledge to a specific decision or action, but rather of a kind of knowing that was inherent in their action. On this ground, he concluded that ‘our knowing is in our action’ (Schon, 1983, p. 49). In a similar vein, but motivated by an interest on the biological roots of learning, Maturana and Varela (1998, p. 27, 29) define knowing as ‘effective action’ and observe that ‘all doing is knowing and all knowing is doing’. Recently, Orlikowski (2002) coined the term knowing in practice to claim that ‘… knowledge and practice are reciprocally constitutive, so much that it does not make sense to talk about one without the other’. A common theme in these works is that ‘knowing’ is continually enacted through people’s everyday activity; it does not exist ‘out there’ (incorporated in external objects, routines, or systems) or ‘in here’ (inscribed in human brains, bodies, or communities).

For the purposes of our current analysis, it is important to assess the means through which such enactment relates to e-learning contexts during social production. Social production is the term used by Y. Benkler to define ‘a new modality of organizing production’. Such a modality is ‘radically decentralized, collaborative, and non-proprietary; based on sharing resources and outputs among widely distributed, loosely connected individuals who cooperate with each other without relying on either market signals or managerial commands’ (Benkler, 2006, p. 60). Today, there are numerous examples illustrating how social production materializes into new and innovative products and services. For instance, projects such as Linux, Apache, Firefox and Wikipedia rely on the social production of open source software (Gilbert & Karahalios, 2009). Social production is also catalyzing traditional industries. For example, tourism is rapidly transformed from a business-to-business or business-to-customer industry to an information-based industry. Collaborative frameworks such as eKoNES (Akoumianakis, 2010) bring social
production to the forefront, changing radically the way in which products are assembled, marketed and used. Similarly, consumer good industries turn into social production processes as a means for sustaining innovation. In this vein, the notion of user toolkits for innovation (von Hippel & Katz, 2002; Franke & von Hippel, 2003) generalizes an approach where innovative user communities are provided with a means (i.e., virtual prototypes, visual artefacts and design abstractions) for articulating new design capabilities.

In all these examples, learning emerges and obtains material forms through the recurrent and socially constituted activities of collaborators as they engage in social production of some type of product. The rationale for such learning results from the heterogeneity of user demand that makes custom, ‘precisely right’ solutions valuable to buyers. As to the type of learning that takes place, it is by and large emergent and unpredictable. Arguably, in none of the examples mentioned above, can the result be foreseen from the start or replicated in exactly the same manner by different configurations of people, artefacts and social relations. Consequently, learning during social production exhibits the features of an emergent knowledge process, conceived of as embodying ‘… deliberations with no best structure or sequence; requirements for knowledge that are complex, distributed across people and evolving dynamically; as well as an actor set that is unpredictable in terms of job roles or prior knowledge’ (Markus et al., 2002).

Facilitating learning as an emergent knowledge process of ‘knowing’ in virtual settings poses several challenges for e-learning contexts. First and foremost, learning is intertwined with designated artefacts of practice. These need not be codified (i.e., documents, tutorials, slide shows) as in the case of instructor-student-course e-learning contexts, but dynamic, socially constituted and evolving. Secondly, learning is enabled or constrained by inscriptions in technology (i.e., APIs) determining not only what is retained from the activities in which collaborators engage in, but also how these remains are related to the artefacts of practice and their history. Finally, the capacity to learn is determined by the e-learning context’s (or system’s) affordances in the sense that possibilities of action at anyone time are only partially given and dependent on the intent of the actors enacting them. These translate to rather distinct requirements for visualizations aimed to serve learning as knowing in practice.

Visualizing Traces of Practice and Online ‘Tells’ to Foster Emergent Knowledge Processes

Visualization is a technology that can be exploited to capture the social-dynamics of knowing in practice. Traditionally, information visualization supports comprehension and analysis of data by individuals using various techniques (Card, Macinlay & Shneiderman, 1999). In more recent efforts the focus shifts to address visualizations in collaborative settings (Bresciani & Eppler, 2009) and knowledge management (Zhang, Zhong, & Zhang, 2010). There have been various efforts in this direction, following slightly different paths. Fisher and Dourish (2004) in their Soylent system describe how collective activity can be revealed and how single-user experience can be reconfigured to support collaborative tasks. In their Many Eyes system, Viegas et al (2007) elaborate on how public visualizations can be created using user-uploaded data and then how such visualizations facilitate large-scale collaborative usage. In CodeSaw, Gilbert and Karahalios (2009) examine social production processes in open source software development communities focusing on visualizations that impact the work of the communities themselves.

A common thrust in these efforts is to capture data that is created (locally) at remote sites and turn it into knowledge shareable and accountable by collaborators. Although this approach may be sufficient to provide informative accounts of what knowledge is articulated across sites, thus answering questions on taxonomic forms of knowledge (i.e., tacit versus explicit), networking patterns and structure, it is not immediately evident how it serves knowing in practice. Specifically, in collaborative settings where the virtual is intertwined with the material realities of collaborators, it is important for visualization to capture not only what knowledge is created or transferred but also emergent properties of knowledge processes i.e., what use knowledge is put into and how this relates to new capabilities.

There are at least two key research challenges in this context. The first is retaining and managing the electronic traces of collaboration in virtual settings. If such traces are retained in an e-learning context then visualization can provide useful insights to past and on-going practice of collaborators, thus to the learning that takes place. In this account, visualization can be conceived of as a tool for excavating virtual settlements, thereby fostering learning. The second challenge stems from the situational nature of learning which necessitates analysis of the intertwining between online and offline activities. Phrased differently, in organizational e-learning contexts learning is not only
about using knowledge to determine practice, but also about understanding how practice re-constructs or creates new knowledge. This time visualization should aim to address the more demanding tasks of amplifying the enactment of knowledge in the collaborators’ local settings and unfolding how such enactments are manifested online to enrich the community’s social capital. Collectively, these challenges motivate a research agenda that expands the potential contributions of visualization to e-learning.

**Distributed Collective Practice Toolkits as e-Learning Contexts for Organizational ‘Knowing’**

Recent collaborative R&D efforts have explored the challenges identified in the previous section in different application domains where social production becomes predominant modality for organizing collaborative work. The focus has been on building distributed collective practice (DCP) toolkits (Akoumianakis, 2009a; Akoumianakis, 2009b) to promote a practice-based perspective on distributed organizing and learning. Before presenting concrete experiences, this section provides an informative review of DCP toolkits, concentrating on their components that rely on visualization to foster learning as knowing in practice.

DCP toolkits are software components intended to facilitate co-engagement of a virtual ensemble in a designated practice. Architecturally, they are characterized by two interoperating constituents – a community management component and a practice vocabulary (Akoumianakis, 2010). Their design is grounded on practice-oriented information systems research (Orlikowski, 2002), thus promoting analysis of collaborative engagements based on technology constituting structures and cultural artefacts of practice. Technology constituting structures are broadly conceived of as functional inscriptions for community management (i.e., registration, communication, information sharing) and non-functional qualities such as connectivity, plasticity and social translucence. Cultural artefacts are traceable evidence of a community’s existence, purpose and underlying practice. A recent study presents an analysis of how these two concepts can be used to classify a broad range of collaborative technologies and social networking sites in terms of their practice orientation (Akoumianakis, 2009b).

In DCP toolkits visualizations are recruited for two primary reasons. The first is to represent elements of a practice vocabulary whose manipulation leads to acts of communication in virtual space. Typically, such visualizations employ metaphor to facilitate sense making and to anchor collaborative activity in virtual settings. The second reason is to consolidate traces of collaboration as retained by the DCP toolkit. This time visualization provides a means for excavating historical remains of collaborative practice within the virtual settlement of a DCP toolkit.

**Visualizations as Practice Vocabularies**

A practice vocabulary may be designed so as to reconstruct elements of an accepted and widely used practice or to enable change by fostering totally new practices. To this end, practice vocabularies implement primitive objects which allow assembly of more complex collaborative artefacts of various types. One category of artefacts depicts the result of individual or collective contributions, e.g., product models, standards, style sheets, etc. Another category of assembled artefacts may detail interdependencies of tasks or objects in a cooperative setting, e.g., organizational charts, classification schemes, taxonomies, etc. Yet another category may comprise artefacts that specify a protocol of interaction in the light of task interdependencies e.g., production processes and workflows, schedules, office procedures, bug report forms. Through these artefacts, toolkits allow graceful translation of the virtual, which is recognizable and accountable by all, to the local setting of different partners and vice versa. Thus, devising appropriate visualizations that afford representation of practice vocabularies serves not only the purpose of transferring knowledge across (physical or field) boundaries, but also the goal of consolidating the collective intelligence of a virtual ensemble.

**Visualizing Online Traces of Collaboration**

The second role of visualization is to represent virtual ‘tells’ of collaborators as they engage in a designated practice. This kind of visualization differs from the previous type as it is primarily intended to either reconstruct an event of the past (through its virtual remains) or to frame on-going activities of a virtual ensemble in relation to an evolving agenda. In effect, such visualizations are intended to capture social data such as the users’ actions on artefacts of
practice, messages exchanged, issues raised, postings in forums, etc. Frequently, such online traces of collaboration depict offline activities. Thus, social proxies are devised and retained to anchor the offline activity in relation to the current state in the virtual setting. These requirements make the design of visualizations more complex to implement as they are strongly dependent on affordances (or non-functional qualities) inscribed into the DCP toolkit, such as the DCP toolkit’s API, type and form of metadata, compliance to de facto standards, etc. In turn, these by and large, determine the social data retained and how they are filtered or transformed to the target data set for visualization.

Case Studies

This section reports on two representative case studies that provide useful insights on how visualization has been used to facilitate ‘knowing’ in practice using DCP toolkits in two different scenarios of social production, namely music notation lessons and assembly of information-based products such as vacation packages for tourists. The music notation lesson case (Akoumianakis & Alexandraki, 2010; Alexandraki & Akoumianakis, 2010) provides insights into virtual learning by reconstructing widely accepted practices based on established music notation constructs. The vacation package assembly case (Akoumianakis, 2010) presents an example where new virtual practices (and their supporting knowledge) are acquired, internalized and enacted by members engaging in cross-organizational collaboration. Despite their difference in focus, both cases emphasize learning that is emergent and framed in the social production of designated collective artefacts. Table 1 summarizes the conditions and criteria of the two cases.

<table>
<thead>
<tr>
<th>Music notation lessons</th>
<th>Community setting</th>
<th>Shared practice</th>
<th>Local settings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moderated squads comprising one tutor and a music learning community of peers</td>
<td>Interpreting and negotiating music scores to acquire new music performance skills</td>
<td>Identical and differentiated only by the performers’ choice of musical instrument</td>
</tr>
<tr>
<td>Cross-organization product development</td>
<td>Virtual alliance of business partners offering competing or complementary services</td>
<td>Negotiating details of aggregate product lines (vacation packages) by assembling services</td>
<td>Differentiated by organizational boundaries and locally instituted work practices</td>
</tr>
</tbody>
</table>

Table 1. Cross-case conditions

Figure 1. The distributed music notation lesson toolkit
Case 1: Music Notation Lessons

A music notation lesson is a multi-site engagement by one moderator (or music theory tutor) and several participants with the aim of constructing, negotiating and reconstructing a music notation so as to produce individual and collaborative music performances. To conduct a music notation lesson, moderators prepare music materials (i.e., notation, recordings, videos, etc), schedule the music lesson and invite participants in synchronous collaborative practice sessions. Participants may access shared contents in their own pace, while during the lesson they negotiate the act of interpreting music materials against their technical virtuosity. In turn, this is manifested as individual or joint music performance. There are two prerequisite for taking part in music notation lessons. Firstly, users must accept the moderator’s invitation and register to the corresponding ‘room’. Such ‘rooms’ are dedicated portlets of the Liferay content management system, which serves as the community support medium. Registration is a two-stage process where participants first become members of the community (by building their music profile) and then register to ‘rooms’. Secondly, users download the practice toolkit, which allows them to co-engage synchronously in the micro-negotiations of a music lesson. An example of this client toolkit is depicted in Figure 1. As shown, the toolkit implements a dedicated room for each participant with online material related to the music notation lesson (i.e., the replicated representation of music) and the shared resources (private space) as extracted from the Liferay room.

The role of visualization in a music notation lesson

Visualization of the designated practice vocabulary amounts to assembling static learning materials available in the Liferay portal and then rendering them through visual encodings that afford interactive manipulation. For instance, static XML scripts of a music agenda can be visualized as an interactive music score using JMusic toolkit. As a result the music score can be replicated across sites and become sensitive to collaborative actions. In effect, the music score bridges across two virtual settlements – the Liferay community where collaboration is asynchronous and the music performance toolkit where collaborative activity is synchronous and dynamic. The music score acts as a boundary object which translates knowledge retained in one virtual settlement as XML script to another virtual settlement. The translation entails much more than the technicalities of mapping elements of a source domain (i.e., XML tags) to symbols in a target presentation vocabulary (Java Swing objects of the interactive music score). It assigns meaning to an unknown context as learners need not and typically do not understand XML. As a result, through translation the scope of the learning object is enhanced since it now affords a range of uses related to or emerging from the actual performance.

A consequence of visualizing the music score is that it is now possible to retain its actual use (i.e., how it is modified and reconstructed) during a music session. Such use can be exploited to enhance our understanding of the collaboration evolving around the music score and to reveal how codified knowledge is put into practice. To this effect, the JMusic toolkit and API were augmented to establish new functionalities, including programmatic manipulation of current object of focus, single and multiple note selection, annotations, marking score segments, changing music parameters, etc. (see Figure 2) so as to establish proxies for what musicians typically do offline. Additional collaborative features were aimed for replicating music scores across sites, granting access and control, fostering awareness and social translucence (i.e., social proxies for current object of focus, participants, current floor holder), managing annotations for conveying intentions as well as visual enhancements of the music score to depict scent and cues.
Limitations, revisions and concluding remarks

A number of experiments have been conducted to assess the value of the DCP toolkit for network music performance (Alexandraki & Akoumianakis, 2010; Akoumianakis & Alexandraki, 2010). Through these experiments it became evident that learning occurs as collaborators negotiate shared resources and artefacts to record a music performance. The learning outcomes include not only improvements in virtuosity or the recordable individual and group music performances, but also the dynamics of collaboration leading to these accomplishments as traced by the toolkit. In fact, these can be reconstructed at a later time to provide useful insights to what has taken place in a virtual ensemble and how a designated cultural artefact of practice (i.e., the recorded music performance) was achieved. It was also revealed that there are aspects of learning which are constrained by the affordances inscribed into the music notation toolkit. Specifically, it was observed that correct musical performance could still occur through suboptimal material actions, such as misuse or misplacement of instruments, wrong notes and erroneous rhythmic patterns, as enacted in practice. It follows that an appropriate e-learning context should be capable of revealing learning not only as outcomes but also as enacted capability. This can be achieved by inscribing new affordances into the toolkit so as to allow designated artefacts of practice to sense local (offline) actions, thus intertwining between online and offline constituents of practice. To respond to these shortcomings, score following capability was introduced through audio to score alignment and multimodal note selection.

Summarizing our experiences, several conclusions can be drawn regarding the relevance of visualization to learning through co-engagement in music notation lessons. Firstly, with respect to the taxonomic classification of knowledge visualization by Eppler and Burkard (2004), our music notation toolkit uses visual metaphors to provide a language for collaboration. The score is interactively manifested to provide a shared context of reference for the collaborators, structuring unknown contexts (i.e., XML) and/or actions and assigning them with meaning. In line with other research in HCI, the visual score and its manipulation do not match precisely the real world metaphor. Affordances enable the visual metaphor to exhibit plasticity, awareness of collaborative engagement and social translucence, thus serving as boundary artefact bridging across online and offline settings. Secondly, learning is not revealed so much from what is readily available as encoded, pre-existent and ready-to-consume knowledge, but from the enacted capabilities acquired in practice through collaborative construction, negotiation and reconstruction of cultural artefacts of practice. Detecting (or excavating) such learning amounts to visualizing online traces of collaborative behaviour and the virtual ensemble’s cultural remains (i.e., recorded performance).

Case 2: Vacation Package Assembly

This case was conceived and designed to assess social production of in-vacation packages for tourists. These are information-based products which act either as supplements to vacation services acquired through conventional means or as catalysts for selecting a destination site. Vacation packages are assembled by electronic squads which in
themselves exhibit an emergent structure. Initially, a squad is dynamically formed by all neighbours registered in the neighbourhoods (i.e., accommodation, food and beverage, transportation, cultural heritage, etc) assigned to a designated package. Although membership is explicitly confirmed, members can decide to opt out at any time. Consequently, a squad may undergo several cycles of reformation until it becomes stable. A vacation package is incrementally constructed and negotiated on the grounds of issues arising in due course. Such issues may range from setting attributes of the package (i.e., duration for certain activities) to negotiating prices or discount policies for service offerings. Once all issues are resolved, the vacation package is assembled into different possible presentation vocabularies (see Figure 3).

The left-hand instance in Figure 3 presents assembled packages in a portlet context. The annotated sequence of actions designates how a package may be explored by customers. The right hand instance presents exactly the same content augmented using social widgets for tagging, rating, commenting, etc. For electronic squads, vacation packages remain aggregate products until prospective customers externalise their intentions by reflecting upon or choosing specific vacation arrangements and business partners. This is a form of tailoring an abstraction so as to translate it into a concrete offering for specific requirements. Through such tailoring, the abstract package may be instantiated into several concrete vacation plans, all within the scope of the package family. Micro-negotiations between squad members are manifested using a different visual vocabulary (see Figure 4). As shown, the package is visualized as a two-dimensional activity panel (or schedule) with individual activities represented through social proxies referred to as elastic buttons (Akoumianakis, 2010). These are socially translucent proxies for designated services, making use of visual scent to encapsulate the history of negotiations (i.e., comments, tags, issues raised, etc) for an activity.

In such a setting, learning as enacted capability at individual, group or alliance level can be revealed through virtual excavation. For instance, at the alliance level, it is possible to assess the types of vacation packages assembled over time, how they are shaped by requirements of certain user groups (i.e., elderly versus young people) and how the alliance responds to these requirements. At the level of electronic squads, enacted capability is revealed through patterns materializing into cyber-structures such as cliques (i.e., consumption of service X correlates with consumption of service Y across several vacation packages). At individual (organizational) level, learning amounts to assessing implications of co-engagement on the business strategy of different types of partners, such as SMEs or large vacation establishments.

Figure 3. Vacation packages assembled and published using different presentation vocabularies
The role of visualization in vacation package assembly

In an effort to explore some of these issues a variety of social visualization mechanisms have been introduced in the vacation assembly toolkit. Figure 5 presents a high level depiction of the practice agenda of an electronic squad (i.e., an elaborated vacation package in the tailoring phase). The network structure depicts the members of the electronic squad (red circles in the inner part), their contributions to the specific package (i.e., green circles depicting services offerings) and the requests by prospective customers (blue ovals in the outer parts of the network). In this visualization, network links have standard semantics and serve to indicate relationships of equal strength. Nevertheless, there have been experiments with alternative visualizations where oval properties (such as size, colour and placement) and link width have been used to reveal additional semantics (Akoumianakis, 2010).

A quick overview indicates not only structural properties of the vacation package, but also relative user preferences and competition per service category (i.e., business partners competing for the same service). As this visualization is
dynamic, it allows assessments across time intervals. Figure 6 reflects a different time instance with more requests posed to business partner 16. Similarly, it is possible to excavate changes in customer behaviour (i.e., choices made by user ‘liferay.com.1152’), squad structural properties (i.e., partners committed to or opting out) and popular service types. Cross vacation package analysis is also useful, as it unfolds recurrent engagement of peers in different practice agendas. Finally, these visualizations can also be augmented with dynamic querying features and filtering mechanisms to allow further excavation. For instance, Figure 7 and Figure 8 depict the state of affairs for business partner 16 at a certain point using clustering and/or social zones.

Figure 6. Aggregate account of activities, business partners and prospective customers

Figure 7. Exploring distributed practices by activity state
Using the explorative tools described above, it became possible to conduct assessment across different vacation packages (Akoumianakis et al., 2011) which have revealed interesting insights to learning. Specifically, it was observed that in virtual alliances partners have different conceptions of the learning that takes place and vary in their capacity to learn and appropriate the benefits of virtual networking. For some, learning amounts to internalizing a new technology that suits business objectives. For others, learning is conceived of as new capabilities and continuous attainment of improvement goals.

As an example of the latter, in one of the alliances investigated the primary beneficiaries were large vacation establishments and innovative SMEs offering differentiated products in alternative tourism. They reported that visualizations allowed them to trace behaviour across vacation packages and identify cross-sector cliques (i.e., customers visiting a cultural site tend to dine in a specific tavern, choose the same accommodation, etc) which led them to refine elements of their virtual presence and strategy. Members of the same alliance also reported on limitations and possible extensions in the current version of the vacation package assembly toolkit. For instance, it was claimed that the toolkit should allow them to share vacation packages in popular social networking sites (Facebook, Linkedin and Tripit) as this would enable establishment of persistent micro-communities amongst customers. Clearly, these findings support the view that learning is an enacted capability which can be revealed through visualizations that do not merely consolidate a virtual settlement, but also augment it by making explicit the enacted configurations of people, artefacts and social relations.

Discussion

In the light of these exemplar scenarios, it is argued that visualization can have a profound impact on organizational learning in virtual settings – it can foster social accomplishment by surfacing traces of recurrent acting and co-engagement in practice, thus a kind of ‘knowing’ in practice. At the core of this finding are some hidden qualities that visualizations need to exhibit. In this section, an attempt is made to briefly elaborate on these qualities and assess their implications for novel organisational e-learning contexts. Hopefully, this will make the contributions of the present work more targeted and explicit.
**Visualizations as Languages for Recurrent Co-Engagement in Boundary Practices**

The relationship between learning and ‘knowing’ in practice has been convincingly addressed in existing scholarships (Lave & Wenger, 1991; Suchman, 2000; Orlikowski, 2002). Our contribution extends these results to online collaborative settings by claiming that visualizations can act as linguistic domains for acting and reflecting upon the history of co-engagement in a virtual practice. The above rests on a conception of visualizations as boundary artefacts fostering learning as enacted capability in communities of practice. Notably, in both case studies presented earlier, meaning and action, at any given time, are mediated by artefacts. It is the interpretive capacity of these artefacts that anchors unknown contexts and assigns meaning to vacation packages and music performance, respectively. In doing so, these artefacts implicate learnt communicative behaviours that exist only within the respective communities of practice. Thus, these artefacts constitute not only community language in the sense that members use them for ‘speaking’ to another, but also a kind of cross-community language since different communities can make sense of and reconstruct the remains of another.

The two cases bring into sight an additional perspective on the boundary role of visualization. This stems from the emphasis on the boundary between virtual and material aspects of different social worlds and the intertwining between the virtual and the material contexts of an activity. For instance, in vacation package assembly reservation for accommodation comprises a definition of accommodation as a package component (online sub-activity recognizable by all) and declaration of commitment to provide such service by members (offline sub-activity in a partner’s information system). However, ‘knowing’ in practice is revealed only through the ‘boundary to local’ translation enacted by customers and the effects this may have on parties involved. Similarly, in the music notation lesson the score becomes de-contextualized offline in the local setting of a musician, making music performance an activity that spans online and offline contexts. Once again, ‘knowing’ in practice emerges through assessment of how an online context (i.e., score) is enacted and translated into music performance. This leads to an interpretation of practice as institutions of distributed activities spanning virtual (online) and material (offline) contexts. Thus, it stands to argue that for visualizations to act as languages for recurrent co-engagement, they should be designed so as to exhibit affordances that foster translations of boundary to local and vice versa.

**Implications**

The case studies presented in this paper raise several implications for developers of organizational e-learning systems. Firstly, it is important that e-learning systems are designed so as to retain use. The vast majority of existing systems are designed with a primary focus on archiving knowledge (i.e., procedures, templates, specifications, etc) and facilitating sharing across and/or reuse by organizational units or parties. This limits their capacity to foster the process of ‘acting’ and ‘knowing’ which embody elements of social, dynamic and situational conduct. Web 2.0 and the plethora of technologies coined as social computing can improve the current situation. For instance, novel APIs make provisions that retain traffic and capture semantic properties of what users actually do online.

Secondly, visualization can be used to appropriate such social data so as to reveal use patterns across time intervals. This makes the choice of visualization a critical design decision. Our work indicates that instead of relying on structured visualizations with proven and accepted use in learning, such as concept maps for example, designers should explore creative options facilitated by 2D and / or 3D visualization toolkits, undertaking the effort to extend or expand them as needed. It is argued that such toolkits offer greater flexibility for articulating social data and more intuitive representations of complex phenomena.

Finally, in cases where visualization relies on virtual ‘tells’ and electronic remains, it is important that it facilitates excavation in virtual e-learning settlements. This bares implications on the design of learning objects and the affordances inscribed in the e-learning system and the visualization. Specifically, in addition to designing learning objects as collections of content items, exercises and assessments, combined based on a single learning objective, designers should focus on qualities whose presence or absence determines how these objects are used in practice. Our case studies reveal that such qualities require provisions for abstraction, social translucence, awareness and plasticity. In turn, the e-learning system should implement inscriptions that allow it to retain use in forms that can be processed to allow an understanding of how organizational knowledge is put into practice as well as how practice re-constructs available knowledge.
Conclusion

Visualization has a potentially important role in organizational learning – that of fostering learning as ‘knowing’. Our analysis builds on two different cases where visualizations are used to reveal learning and how it emerges in different configurations of people, artefacts and social relations. Both cases depict learning contexts that are detached from the traditional instructor-student-course model of learning that focuses on acquiring knowledge. Instead, they assume learning as social accomplishment, an enacted capability mediated by artefacts, and as such, facilitated by appropriate knowledge visualizations making use of metaphor to visualize domain-oriented concepts and electronic traces of collaborators. Augmentations of the visualizations make provisions for social awareness, plasticity and social translucence to facilitate boundary crossing (in each case) and the intertwining between online and offline practice.

Our main conclusions are three-fold. Firstly, visualization should aim to catalyze not only the transfer of knowledge that is ‘out there’ so that it becomes internalized and codified by the learner, but also what social practice and collaborative doing make this knowledge to be. Secondly, the power of visualization in organizational learning stems as much from symbolic mediation of organizational action as from the capacity to represent intertwined online and offline realities across different boundaries. Finally, visualization conceived of as the organizational language that mediates collaborative activity and leads to recurrent co-engagement in practice, need not be approached through the lens of re-producing online what people are used to doing offline. A more challenging option is to design for totally new practices that sustain novel (learnt) communicative behaviours and induce innovative changes of habit.

References


Screen-capturing System with Two-layer Display for PowerPoint Presentation to Enhance Classroom Education

Yen-Shou Lai, Hung-Hsu Tsai and Pao-Ta Yu

Department of Information Management, National Formosa University, Huwei, Yulin, Taiwan // 1Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan // layysmail@gmail.com.tw // thh@nfu.edu.tw // csipty@cs.ccu.edu.tw

ABSTRACT

This paper proposes a new presentation system integrating a Microsoft PowerPoint presentation in a two-layer method, called the TL system, to promote learning in a physical classroom. With the TL system, teachers can readily control hints or annotations as a way of making them visible or invisible to students so as to reduce information load. In addition, the system can synchronize a teacher’s lecturing actions for a PowerPoint presentation with his/her voice and also create web-based multimedia materials. By reviewing these multimedia materials, students can recall the problem-solving skills the teacher taught them in class. In this study, an exploratory test was carried out with 12 teachers and 91 students in a college of technology. Research data were collected through the instructional technology use questionnaire (ITUQ) for teachers as well as through the course attitudes scale (CAS) and the mathematics achievement test (MAT) for students. The results revealed that the TL system helps teachers to have a smooth teaching process of presenting teaching materials. The students in the experimental group had a better course attitude and achieved higher MAT scores than did those in the conventional group.

Keywords

Classroom learning, Media in education, Multimedia, PowerPoint presentation

Introduction

In traditional classrooms, teachers write learning content on a blackboard at the front of the classroom. What the teacher writes on the blackboard has an exemplary effect on the learners (Havens, 1989; Zhang & Deng, 2005). It is a real-time and flexible process of transferring the knowledge of the learning content from the teacher to the students (Lightfoot, 2005). One advantage of this method is that students can pay more attention to enhance their classroom experiences (Zhang & Deng, 2005). In the traditional classrooms, teachers write the problem-solving procedure on a blackboard in a step-by-step fashion. This allows teachers to emphasize abstract concepts such as switches and branch points in a flexible manner during the instruction. It allows students to feel the subtle rhythms in the teacher’s teaching, and makes it easier for students to comprehend and understand the mathematics learning content (Havens, 1989).

In the teaching process, one drawback of using blackboard writing is the fact that presenting the materials is time-consuming. This is especially true when the teacher’s back obscures a portion of what is written on the blackboard, which increases the level of restiveness in the class. Also, using a backboard this way limits the way of presenting teaching materials. For example, writing on a blackboard provides an inflexible presentation with few colors and styles, as well as difficulties in displaying pictures or multimedia content. Another disadvantage of using a blackboard is that the materials cannot be stored, reused, reproduced, or interchanged. Consequently, students have to make notes or copy the material from the blackboard, all of which is inconvenient (Apperson, Laws, & Scepansky, 2008). This results in an increased external cognitive load and splits the student’s attention. Blackboard writing is limited by the fact that the writing differs from teacher to teacher. In addition, a large amount of chalk dust is produced when using a blackboard.

Recently, many teachers have started to use technological tools to create teaching materials in multimedia formats. It is helpful to effectively scaffold learners (Lai, Tsai, & Yu, 2009; Lightfoot, 2005). However, it is not easy to use technological skills for creating multimedia teaching material. A popular lecturing method is to directly project Microsoft PowerPoint slides from a computer onto a screen (Rankin & Hoaas, 2001). Many studies have indicated that college instructors accompany their lectures with PowerPoint presentations so that their lectures will have a positive effect on their students’ attitude and belief of self-efficacy (Rankin & Hoaas, 2001; Susskind, 2008). Students attending classes where the teacher used PowerPoint believed that the lectures were more organized, clear, and interesting. The PowerPoint presentation can overcome the limits and the disadvantages mentioned for using a blackboard. This is due to the following. First, lecturing using PowerPoint presentation can save writing time in class
because the PowerPoint slides are prepared beforehand. As a result, teachers have more time to interact with their students and effectively proceed to instructional activities. Second, while using PowerPoint presentations, teachers face students and not the blackboard. Thus, the teacher’s back will not obscure the student’s sight. Third, PowerPoint slides can be quickly stored, reused, reproduced, and interchanged. The contents of PowerPoint slides can be produced in various formats such as texts, tables, pictures, graphs, sounds, visual data, video clips, and so on. Lecturing with PowerPoint presentation can be used to provide important outlines, key terms or phrases, different background formatting, and various formats of annotation (Apperson, Laws, & Scepansky, 2008). Unlike with blackboard writing, teachers can present integrated multimedia instructions, including media format selection, and have random access to multimedia instruction (Corbeil, 2007). At the same time, teachers can easily convert PowerPoint content to multimedia formats. Finally, lecturing using PowerPoint presentations can enhance the lecturer’s ability to order and pace his/her lecture and present a clear summary (Lowry, 1999). This is because lecturers can easily control the lecture content and display the sequences when using PowerPoint slides.

However, PowerPoint presentations still have some potential limits. Each slide in the PowerPoint format contains only a small amount of information. In addition, a PowerPoint presentation is similar to a bullet-style presentation and is only suitable for a low level of information transfer (Tufte, 2003). Lecturing using PowerPoint results in a weak analysis of the learning content (Gabriel, 2008; Zhang, Zhao, & Nunamaker, 2004). In addition, when students review the PowerPoint material before or after class, the review is not effective because the material contains only key terms and pictures, but is not accompanied by the teacher’s action and voice. One solution to overcome the above problems is to adopt screen-capturing software that can create a video file of the teacher’s voice and the entire PowerPoint presentation. The most widely used software applications include Camtasia Studio, PowerCam, Articulate Presenter, etc. Any teacher can easily produce the video learning materials for a cyber classroom (Chen & Wang, 2008), and it provides students with more learning opportunities.

At present, teachers use screen-capturing software for delivering multimedia teaching contents via the Internet. However, these software applications lack one important function. In the TL system, a teacher, when lecturing, can see some of the teaching materials while those same materials are invisible to the students. From the student’s point of view, this function is helpful to reduce the extraneous cognitive load because all learning content will not appear at once, and those materials that do not need to appear do not. From a teacher’s point of view, this function is helpful to reduce tension for the student because the teacher can see the hints and annotations while teaching even though they remain invisible to students. In addition, this function can make the lecturing smoother, because the teacher can use the auxiliary descriptions or notes as reminders. At present, teachers can make speaking notes and reminders in the Note Page View for the PowerPoint slides. Unfortunately, these notes are only written in a text box using a simple text format. The limitations of using Note Page View are that the text box is locked at a fixed position in the PowerPoint slide and speaking notes cannot be made during a presentation.

This paper proposes a screen-capturing system with a two-layer display for PowerPoint presentations, called the TL system. The summarization of this study is threefold: (a) develop the TL system to record the slide content and any lecturing notes of the teacher in his/her handwriting and voice, and then upload the recorded files to the LMS; (b) examine whether the TL system is helpful to teachers during instruction in class; and (c) investigate whether there are any differences in the learning effects of the students between using PowerPoint and the TL system.

The rest of this paper is organized as follows. First, we present a literature review and the learning environment involving the TL system. Next, we compare the features of the TL system with those of other developed recording tools and discuss the experimental results. Finally, we present our concluding remarks.

**Literature review**

**Instructional design with cognitive theory**

Multimedia instruction presents multiple materials that are intended to foster learning, including speech, printed text, static graphic, animation, and video (Moreno & Mayer, 2007). The cognitive theory of multimedia learning provides empirical guidelines to promote instructional design so as to achieve meaningful learning (Mayer, 2001). The theory is based on the assumptions of dual channel, limited capacity, and active processing (Moreno & Mayer, 2000). Cognitive load is related to human information-processing capacity (Guttormsen, Schär, & Zimmermann, 2007). When students control the information flow, they can chunk the information in a meaningful way, according to their

**PowerPoint presentation in instruction**

Humans can simultaneously process information coming from auditory stimulus and visual stimulus (Moreno & Mayer, 2000). Recently, face-to-face lectures have been delivered using presentation software such as Microsoft PowerPoint. Using PowerPoint to present multimedia materials in class benefits students (Apperson, Laws, & Scepansky, 2008). PowerPoint slides contain various multimedia formats such as text, chart, graph, sound, and video. The way of lecturing with PowerPoint can provide students with a brief description for teaching sequence and organization of the learning contents. This teaching manner contributes to students’ further constructing learning concepts and conducting information-processing analysis (Susskind, 2008). Atkinson and Mayer (2004) suggested that there are five principles of instructional design to help reduce extraneous cognitive load in PowerPoint: writing a clear headline that explains the main idea of every slide, breaking up the story into digestible bites in the slide sorter view, reducing visual load such as moving text off-screen and narrating the content, using words with visuals instead of words alone, and rigorously removing every element that does not support the main idea.

**Instructional environment and implementation**

**Instructional environment**

Teachers can use the TL system to construct a new instructional environment with dual-channel display for classroom learning. Figure 1 depicts the pictorial structure of the new instructional environment, which, in general, consists of a PC (notebook or tablet PC), two (or one) projectors, a projection screen, and the TL system. The environment can help teachers to have more flexible and useful presentations, especially when using PowerPoint.

![Figure 1. The structure of the instructional environment](image)

**System design and structure**

The TL system records the multimedia version of teaching materials during instruction process. The materials include slide content accompanied by the teacher’s voice and all tracks left on the slides (note that the tracks probably consist of teacher’s handwriting, the cursor’s movement, and the display and manipulation of learning objects). Therefore, students can review the lectures by studying the multimedia teaching materials after class.
The main feature of the TL system, which differs from other existing systems, as shown in Table 1, is that some auxiliary materials or hints are visible to teachers but invisible to students. The feature is devised by using a two-layer display technique (Aboelsaadat & Balakrishnan, 2004), as shown in Figure 2. The two-layer display consists of the presentation layer and the script layer. The presentation layer includes the display items, which are teaching objects visible to teachers and students, and possibly contains texts, images, sounds, and movies. The script layer includes the hint items, which are auxiliary semi-transparent objects for teachers that are invisible to students, and possibly contains guidelines and scripts. When the teacher needs to display these invisible objects in the script layer, the teacher moves them to the presentation layer. A case of displaying method is that teacher can draw the invisible scripts of the script layer on the presentation layer. The aim of using the two-layer display is to help the teacher to have teaching guidelines while lecturing. Moreover, the teacher uses the feature in lecturing to make the teaching process go more smoothly and to reduce the teacher’s anxiety. It also reduces students’ extraneous cognitive load because not all of the content in a slide will appear (Mayer & Moreno, 2003).

Table 1. Comparison of screen-capturing software

<table>
<thead>
<tr>
<th>Items</th>
<th>Tools</th>
<th>Screen-capturing software</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Articulate Presenter</td>
<td>Camtasia Studio</td>
</tr>
<tr>
<td>Recording method</td>
<td>PowerPoint recording</td>
<td>Screen recording</td>
</tr>
<tr>
<td>Supports post-production</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Records voice, monitors, and</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>cursor tracks synchronously</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials index with slide</td>
<td>No</td>
<td>Medium (Post-production)</td>
</tr>
<tr>
<td>Summarizes theme of slides</td>
<td>Manual</td>
<td>Medium (semiautomatic)</td>
</tr>
<tr>
<td>in table for indexing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selects monitor dpi freely</td>
<td>Multi-selection</td>
<td>Multi-selection</td>
</tr>
<tr>
<td>Support two-layer display</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>mechanism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports multi-monitor</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>recording and displaying</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 shows the five principle components of the TL system: project, layer, multi-display, recording, and management. The project component creates, opens, saves, or closes a project and imports PowerPoint files. The layer component implements the script layer and the presentation layer. The multi-display component supports a multi-monitor display environment. The teacher can easily extend the learning material or annotations to the other monitors. The recording component records audio, handwriting, cursor movement, slide content, and more. The management component builds tree structures to manage project material. It offers the teacher a friendly interface to test the system, draw grids on the screen, set up the recording frame per second, and export the streaming video file. In the TL system, teaching materials are organized as a tree structure, as shown in Figure 4. The teacher can easily maintain the materials. Students can linearly or randomly access the multimedia materials with webpages via the
Internet, as shown in Figure 5, which can effectively help students to review those teaching materials. The reason is that students can listen to the verbal presentation of teaching materials as they view the visual presentation, instead of merely viewing static PowerPoint files or documents or listening to a recording of the teacher’s voice after class.

**Figure 3.** The structure of the TL system

**Figure 4.** Teacher’s authoring mode

**Figure 5.** Display mode for students on the Internet
Scenario of instructional implementation

This teaching scenario demonstrates that the teacher applies the TL system to an instructional procedure. The scenario is to teach a unit of calculus, in which students find the volume of the solid obtained by rotating a certain curve line about the x-axis. Figure 6 displays several screenshots of the teacher’s notebook computer. The teacher can see all objects in the slide, but some learning objects in the slide are invisible to students (as shown in Figure 7a).

Figure 6. The teacher’s screen in class

Figure 7. Six snapshots in the teaching process
The teaching procedure of the scenario is described as follows:

1. **Review the background for the teaching units.** The teacher presents the PowerPoint slide showing only the $x$-axis, $y$-axis, and a curve $f(x)$ from $a$ to $b$ (Figure 7a). In the first five to ten minutes of class, the teacher helps students review the background for the teaching units, such as how to compute the area of the curve line to the $x$-axis. The teacher lectures the concepts of the key points, “slice” and “sum.”

2. **Present the new teaching materials.** Figures 7b to 7e show four snapshots of the teaching process. Figure 7b illustrates the initial content of the slide while the teacher is teaching. Students see only the initial content of the slide, but the teacher can see the complete content, as in Figure 6. Figure 7c shows that the teacher uses a digital pen to sketch the cylinder. The teacher draws the sketch and slices the solid, $S$, into pieces by following the annotated hints. Figure 7d shows that the teacher has written the keywords and explanation. While teaching, the teacher can refer to the guidelines in the script layer to present topics such as slicing the cylinders like a loaf of bread. The purpose of showing auxiliary hints is to remind the teacher to emphasize the skill of solving a problem, such as getting the change of quantity, from $a$ to $b$, instead of making the error from $b$ to $a$. Figure 7e shows the teacher’s handwritten equations. The first equation illustrates the area of each piece of a cylinder. The remaining equations represent the summation of the cylinders to get the volume.

3. **Review the key points of the teaching materials.** The teacher concludes and reviews the problem-solving skills, including how to sketch the solid, how to find the formula $f(x)$, how to find the limits of integration, and how to integrate the area (as shown in Figure 7f).

**Method**

**Participants**

A quasi-experiment was conducted to investigate the effects of applying the TL system in the physical classroom. Twelve teachers in a technology college of a university in middle Taiwan participated in the trial of the system. The teachers have experiences in using or producing PowerPoint slides as teaching materials. Moreover, 91 sophomore and junior students in the same technology college (55 men and 36 women, with a mean age of 19.8 years), who selected the linear algebra course enrolled in the experiment. The students were randomly assigned to two classes after they registered for the course. Then, the two classes were randomly assigned as the experimental group and the control group. The experimental group (42 students) was lectured by a teacher using the TL system, and the control group (49 students) was lectured by a teacher using a PowerPoint presentation.

The internal controls, which included teaching hours, instruments, course content, and instructor, were the same for the two groups. The experiment ran for four months, from September 2008 to December 2008. Each class met three times per week, for 50 minutes each session. The external controls included random assignment and analysis of covariance. First, at the beginning of the semester, the students were randomly assigned to each class. Second, the mathematics achievement pretest score was taken as a covariance variable so as to understand the students’ background regarding the learning content. The time factor, the course period of the experiment, is a history event that occurred outside of the experimental situation but may have affected the internal validity. It may also have affected the participants’ responses to experimental procedures. For example, some of the students may have gained knowledge related to the teaching materials outside of the classrooms through extra classes rather than solely using the TL system. Although the quasi-experimental design may be weaker than the experimental design, it is still used for much research in social sciences (Ainsworth, 2007; Kwon & Cifuentes, 2007; Lee, Shen, & Tsai, 2008; Wei, Chen, Kinshuk, & Hsu, 2009).

**Instruments**

*Instructional technology use questionnaire (ITUQ) for teachers*

In this paper, the ITUQ for teachers was investigated by using the questionnaire, which includes perceived use and trial experience of the TL system. Perceived use is the degree to which teachers believe that using a particular system would be easy and enhance their performance. Here, the ITUQ was modified and edited by the questionnaire of Davis’s study (1993) and the teachers’ trial experience (Hwang, Tsai, Tsai, & Tseng, 2008). There were 8 items in
the ITUQ, as shown in Table 2. Each item was assessed on a five-point Likert scale that ranged from 1 (strongly disagree) to 5 (strongly agree). The Cronbach’s alpha reliability coefficient for the questionnaire was 0.82.

Course attitudes scale (CAS) for students

We collected data from a 21-item CAS that students completed at the end of the semester. The scale included 6 items used by Apperson, Laws, and Scepansky (2008) to evaluate students’ attitudes toward the course, as well as 15 items from Susskind (2008) for material presented in the classroom via PowerPoint. Each item was measured on a seven-point Likert scale where 1 and 7 indicated strongly disagree and strongly agree, respectively. Reliability for the scale was measured with Cronbach’s alpha, which yielded a score of 0.85.

Mathematics achievement test (MAT)

A 33-item test for linear algebra was designed for pretest and posttest. The test included three components. The first component, which included seven items (28 points), referred to the problem-solving of procedural tasks that assess students’ ability to solve standard tasks that demand basic skills. Therefore, the purpose of the component was to evaluate students’ basic problem-solving skills. The second six-item component (30 points) referred to the problem-solving of transfer tasks. These items were more complex than those in the first component. The aim of the second component was to assess whether students had higher level skills. The third six-item (42-point) component referred to providing mathematical explanations. The mathematical-explanation items were identical in the pretest and posttest, whereas, the transfer tasks and the standard tasks were different.

Procedure

Twelve teachers and 91 sophomore and junior students in two classes contributed to the trial of the system. The classes agreed to teach with the instructional technology to support teaching of their lectures in the autumn 2008 semester. The control group, without using the TL system, received a lecture accompanied by a PowerPoint presentation. The experimental group used the TL system. A main difference between the two is that some content is invisible to students but visible to teachers in the TL system. The teaching materials were uploaded to the LMS for students. Students in the control group were able to download the PowerPoint presentation after class, and students in the experimental group were able to download the audio-video materials produced by the TL system after class. Absent students were also able to study the teaching materials by using the LMS. The procedure of the TL group, as shown in Figure 8, is described as follows.

1. Pretest: At the beginning of the semester, the students were asked to take the mathematics achievement test (MAT) to understand their background about the learning content. A change between the pretest and posttest is that teachers integrate the TL system into their instructional methods.
2. Before lecturing: Teachers create a new project of the TL system and import PowerPoint files into the project. Then, teachers use the TL system to edit the hint items and annotations and then arrange the presentation sequence of slide objects.
3. During lecturing: Teachers present the learning content and use the TL system to record their lecturing actions and voice simultaneously.
4. After class: Teachers post-edit the streaming video to optimize the materials and then upload the recorded materials in audio-video format to the LMS for students.
5. Posttest: At the end of the semester, teachers fill out the ITUQ. Students were asked to fill out the CAS and to do the MAT.

Data analysis

Research data were collected through the ITUQ for teachers and through the CAS and the MAT for students. Statistical analyses (t-test and ANCOVA) were conducted to investigate the differences between the experimental group and the control group related to questions, and learning achievement and effect sizes (Cohen’s d) were calculated. Here, the significant level was set at $p = 0.05$.

Results

Teachers’ data for ITUQ

The investigation of teachers’ data for instructional technology use aimed at introducing the effectiveness of the TL system to the teachers who were the users of the system. Table 2 shows the statistical results of the questionnaire, and some qualitative comments that the teachers made about the TL system.

It was found that over 67% of the teachers agreed that the TL system can make better recorded learning materials than the DV camera ($M = 3.67, SD = 1.15$). About 25% of the teachers disagreed with this statement possibly because they had to prepare more materials using the TL system rather than the DV camera to teach. Nearly 67% of the teachers agreed that the interface of the TL system was user-friendly ($M = 3.75, SD = 0.87$). Some of them suggested adding illustration to engage users. Only one teacher was not good at using the interface of the system.

Nearly 84% of the teachers agreed that the TL system was helpful to them ($M = 4.08, SD = 0.90$). Over 92% of the teachers agreed that viewing the process with the recorded materials helps them to revise their instructions ($M = 4.17, SD = 0.58$). About 67% of the teachers agreed that the function that some objects were visible to teachers but invisible to students can make the teaching process go more smoothly ($M = 3.83, SD = 0.72$). Nearly 58% of the teachers agreed that having some objects visible to teachers but invisible to students can reduce anxiety while teaching ($M = 3.42, SD = 1.44$). About 25% of the teachers disagreed with this item possibly because they had more than five years of teaching experience.

Moreover, 75% of the teachers would like to employ the TL system in the future ($M = 3.92, SD = 0.90$), and 75% would recommend it to other teachers ($M = 3.92, SD = 0.90$). Again, one of the 12 teachers hesitated to use the TL system because s/he preferred to use the blackboard rather than computer while teaching. We can conclude that the TL system is acceptable for most of teachers.

Table 2. Percentage, means ($M$), and standard deviation ($SD$) of ITUQ

<table>
<thead>
<tr>
<th>Question (5 = strongly agree; 1 = strongly disagree)</th>
<th>SA (%)</th>
<th>AG (%)</th>
<th>N (%)</th>
<th>DA (%)</th>
<th>SD (%)</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I will use the TL system to make the recorded</td>
<td>25</td>
<td>42</td>
<td>8</td>
<td>25</td>
<td>0</td>
<td>3.67</td>
</tr>
<tr>
<td>learning materials rather than using a DV camera.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.15)</td>
</tr>
<tr>
<td>The interface of the TL system is user-friendly.</td>
<td>17</td>
<td>50</td>
<td>25</td>
<td>8</td>
<td>0</td>
<td>3.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.87)</td>
</tr>
<tr>
<td>The TL system was helpful to me.</td>
<td>34</td>
<td>50</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>4.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.90)</td>
</tr>
<tr>
<td>Viewing the recorded materials helps me to revise</td>
<td>25</td>
<td>67</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4.17</td>
</tr>
<tr>
<td>my teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.58)</td>
</tr>
<tr>
<td>While teaching, having some objects visible to</td>
<td>17</td>
<td>50</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.15)</td>
</tr>
</tbody>
</table>
teachers but invisible to students can make my teaching process smoother. While teaching, the function that some objects are visible to teachers but invisible to students can reduce my anxiety. I am willing to use the TL system in my class. I will recommend the TL system to other teachers.

Strongly agree: SG; Agree: AG; Neutral: N; Disagree: DA; Strongly Disagree: SD

Students’ data for CAS

The analyses of independent t-test and the effect size (Cohen’s d) was conducted to assess the effects of the experimental group with the TL system and the control group with PowerPoint on the 21 survey items related to the CAS. Table 3 shows the significant results of the experimental analysis, which indicate that the students in the experimental group felt the instructor’s use of the instructional technology helped them pay attention in class, t (89) = 3.170, p < 0.01, d = 0.68. It may be that the teacher effectively presented teaching materials. When the students studied the materials through the LMS platform, they felt that they could recall classroom experiences, t (89) = 4.215, p < 0.001, d = 0.90. The materials on the LMS platform could help students understand the learning content, t (89) = 3.698, p < 0.001, d = 0.78. Students felt they were more confident for the exams, t (89) = 2.024, p < 0.05, d = 0.43. Compared to the experimental group, students in the conventional group felt that they required more reading, t (89) = 3.757, p < 0.001, d = −0.78. They found it difficult to understand the course, t (89) = −2.151, p < 0.05, d = −0.45. They had to work hard, t (89) = −3.010, p < 0.01, d = −0.63. During lectures, they took more notes, t (89) = −2.454, p < 0.05, d = −0.52. They spent more time studying for exams, t (89) = −2.276, p < 0.05, d = −0.48. Accordingly, these results reflected that the TL system was helpful for students in the experimental group.

Table 3. Means (M), standard deviations (SD) and t-test of students’ CAS

<table>
<thead>
<tr>
<th>Question (7, strongly agree; 1, strongly disagree)</th>
<th>Experimental group (n = 42)</th>
<th>Conventional group (n = 49)</th>
<th>t value</th>
<th>Effect size (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compared to other courses, this course required more reading.</td>
<td>5.76(1.05)</td>
<td>6.51(0.84)</td>
<td>−3.757&lt;sup&gt;c&lt;/sup&gt;</td>
<td>−0.78</td>
</tr>
<tr>
<td>Compared to other courses, this course required more writing.</td>
<td>5.47(1.17)</td>
<td>5.91(1.20)</td>
<td>−1.766</td>
<td></td>
</tr>
<tr>
<td>Compared to other courses, this course required doing additional work.</td>
<td>3.86(1.60)</td>
<td>3.84(1.52)</td>
<td>−0.062</td>
<td></td>
</tr>
<tr>
<td>Compared to other courses, this course was difficult to understand.</td>
<td>4.97(1.26)</td>
<td>5.59(1.44)</td>
<td>−2.151&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.45</td>
</tr>
<tr>
<td>Compared to other courses, this course required me to work harder.</td>
<td>5.40(1.15)</td>
<td>6.08(1.00)</td>
<td>−3.010&lt;sup&gt;b&lt;/sup&gt;</td>
<td>−0.63</td>
</tr>
<tr>
<td>I can easily discuss the lecture with classmates afterwards.</td>
<td>4.52(1.09)</td>
<td>4.86(1.24)</td>
<td>−1.351</td>
<td></td>
</tr>
<tr>
<td>The use of instructional technology helps me pay attention in class.</td>
<td>4.81(0.89)</td>
<td>4.04(1.33)</td>
<td>3.170&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.68</td>
</tr>
<tr>
<td>I thought the teacher’s use of instructional technology while teaching was effective.</td>
<td>4.90(0.96)</td>
<td>4.38(1.37)</td>
<td>2.065&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43</td>
</tr>
<tr>
<td>When I studied the material through the LMS platform, I could clearly recall the classroom experience.</td>
<td>5.05(0.83)</td>
<td>4.16(1.12)</td>
<td>4.215&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.90</td>
</tr>
<tr>
<td>The materials on the LMS platform help me understand the learning content.</td>
<td>4.95(0.96)</td>
<td>4.08(1.24)</td>
<td>3.698&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.78</td>
</tr>
<tr>
<td>During the lecture, I took more notes.</td>
<td>4.38(1.17)</td>
<td>5.06(1.43)</td>
<td>−2.454&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.52</td>
</tr>
<tr>
<td>My notes were more useful for exams.</td>
<td>4.17(0.96)</td>
<td>4.28(1.24)</td>
<td>−0.505</td>
<td></td>
</tr>
<tr>
<td>I spent more time studying for exams.</td>
<td>5.17(1.18)</td>
<td>5.73(1.18)</td>
<td>−2.276&lt;sup&gt;a&lt;/sup&gt;</td>
<td>−0.48</td>
</tr>
<tr>
<td>I was more confident about the exams.</td>
<td>3.45(1.21)</td>
<td>2.92(1.28)</td>
<td>2.024&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.43</td>
</tr>
</tbody>
</table>

<sup>a</sup>p<.05, <sup>b</sup>p<.01, <sup>c</sup>p<.001
Student score

Table 4 shows the mean (M), adjusted means, and standard deviations (SD) of students’ MAT scores. A pretest was conducted, which showed that the students in both the experimental group and the control group have equivalent prior knowledge for learning the linear algebra course, \(t(89) = -1.085, p > 0.05\). A one-way analysis of covariance (ANCOVA) was conducted on the scores of the learning goal with the pretest scores used as a covariant. ANCOVA was performed after confirming requirement of homogeneity of within-cell regressions \((F(1, 87) = 0.714; p > 0.05)\). Results of the ANCOVA revealed a statistically significant difference for learning achievement, \(F(1, 88) = 4.22, p < .05\). Effect size indicated that at the end of the study the experimental group outperformed the conventional group in problem-solving tasks \((d = 0.41)\). This indicates that students in the experimental group gained higher scores than those in the control group.

Table 4. Means (M), adjusted means, and standard deviations (SD) of the MAT score

<table>
<thead>
<tr>
<th></th>
<th>Experimental group ((n = 42))</th>
<th>Control group ((n = 49))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
</tr>
<tr>
<td>Pretest</td>
<td>34.48</td>
<td>17.10</td>
</tr>
<tr>
<td>Posttest</td>
<td>69.76</td>
<td>16.01</td>
</tr>
<tr>
<td>Adjusted means</td>
<td>70.65</td>
<td>62.95</td>
</tr>
</tbody>
</table>

Discussion

It should be useful for teachers and learners to apply the TL system in instruction. In learning activities, teachers and students can benefit by following the four findings found in the experimental results. First, teachers indicated that the TL system is helpful for them (as presented in Table 2). An outstandingly useful feature of the system is that, while lecturing, teachers can display some objects of the PowerPoint slides in their own monitor yet prevent the students from viewing them (as shown in Figure 7). Teachers can see the hint items (or annotations) that they made prior to the lecture, which enables them to confidently present materials. These hints help reduce teachers’ anxiety and make the teaching process go more smoothly.

Second, this feature can be integrated with handwriting and pop-up objects to become a new presentation method. The displayed content is semi-transparent for teachers but invisible to students (as illustrated in Figure 2). Teachers can handwrite the content by drawing the strokes on the semi-transparent content on PowerPoint, which is invisible to students. Alternatively, a simple way to show the semi-transparent content is by using the pop-up functions provided by PowerPoint. Several studies indicated that instructors should no longer rely on the traditional blackboard when teaching (Bruun, 2009; Corbeil, 2007; Havens, 1989). It is impossible for the TL system to completely replace the blackboard because teachers may prefer to handwrite information on the blackboard instead of on a computer screen. However, students are interested in studying through teaching materials such as texts, pictures, graphs, tables, charts, and animation (Liu, Liao, & Pratt, 2009). Furthermore, teachers can review the recorded materials and then improve their teaching (as shown in Figure 8).

Third, students felt that the TL system helped them pay attention in class (as presented in Table 3). With the TL system, while teachers are writing, students have the advantage of having a face-to-face interaction with their teacher. The possible reason is that the handwriting process forced students to think critically and enrich their learning experience at a deeper level (Zhang & Deng, 2005). Thus, the teacher’s writing allowed students to pay more attention, especially when learning mathematics, in which there are many problem-solving steps. Students in the conventional group felt that the learning content was difficult. Moreover, they felt it necessary to read more and study harder. The finding is similar to Susskind’s (2008) finding that PowerPoint lectures did not enhance the students’ performance on exams. However, there is no significant difference between two groups in time spent preparing and studying for this class each week. The possible reason is that, although PowerPoint presentations were well-constructed, the typical organizational benefits of lectures might have been minimized through the use of PowerPoint because the slides present only outlines and key points, as opposed to detailed content. Moreover, teachers often overemphasize the items in slides and directly present slides one by one. As a result, students are not encouraged to think deeply about the presented material. The result does not promote student interactivity, engagement with the content, or learning, which is similar to that of the traditional PowerPoint working methods (Zhang, Zhao, & Nunamaker, 2004). On the contrary, PowerPoint presentations gradually reduce the students’
interest, which can lead to confusion and frustration. The teaching manner with the TL system keeps students attentive while teachers handwrite lessons rather than mechanically displaying content from the PowerPoint slides.

Finally, students agreed the materials on the LMS platform can help them understand the learning content (as presented in Table 3). It may be that students can repeatedly study the recorded materials to review teachers’ problem-solving skills and thinking methods (as shown in Figure 5). Accordingly, they easily followed the teachers’ logic because the multimedia teaching materials included imitating teachers’ problem-solving skills, listening to what the teachers said, and reading what the teachers wrote. According to schema acquisition and the borrowing and reorganizing principles of cognitive load theory (Guttormsen, Schär, & Zimmermann, 2007), the method of studying the recorded materials can benefit students because they can primarily build the bulk of information in long-term memory (Leahy & Sweller, 2008). When students in the experimental group studied the teaching materials, they had a feeling that they were in class. In other words, they can almost recall learning activities in the class situation. This is helpful to transform the knowledge of the teacher’s problem-solving skills to the students. Moreover, teachers can use the TL system to produce multimedia teaching content with webpages through the Internet (as shown in Figure 4). Then, the content can be conveniently uploaded in the LMS system or web-based platform for learning by Web 2.0 to provide students with interactive functions in the cyber classroom (Chen & Wang, 2008). This provides learners with more opportunities to enhance the learning process.

Conclusion

This paper has proposed the TL system. The survey and empirical measures appear to measure that the TL system is helpful to teachers and the students in traditional classrooms. While the teacher lectures, some items in the PowerPoint slides displayed in the teacher’s monitor (or notebook) are visible to teachers but invisible to students. This function is helpful for teachers because it allows them to control the presentation sequence of the PowerPoint slides. Teachers can conveniently refer to the hint items or annotations to smooth their presentation. The TL system is also helpful for improving students’ learning effects while requiring teachers’ handwriting for solving problems (Bruun, 2009). Additionally, teachers’ lecturing voices and whole teaching actions on PowerPoint slides are simultaneously recorded into multimedia teaching content in a web-based format. Students, therefore, can recall teaching activities such as teachers’ problem-solving skills and logic. Finally, the results show that students in the experimental group had a more positive attitude toward learning and gained higher mathematics scores than students in the control group. The TL system benefits students while they review teaching content asynchronously in the cyber classroom (Chen & Wang, 2008).

Acknowledgements

We thank the reviewers for their comments, which made the paper better and more readable, as well as the National Council of Taiwan NSC 96-2520-S-194-002-MY3 and NSC 96-2221-E-150-062.

References


Metacognitive Process in Online Text Construction

Hui-Chin Yeh and Yu-Fen Yang
Graduate school of applied foreign languages, National Yunlin University of Science & Technology, Douliu, Yunlin, Taiwan // hyeh@yuntech.edu.tw // yangy@yuntech.edu.tw

ABSTRACT
Many studies have reported the results of how students constructed texts, yet the cognitive process of how texts were constructed by sentences and how the constructive process was formulated and reformulated have rarely been addressed. This study aims to develop a computer-aided text construction system which supports 83 English as a Foreign Language (EFL) university students in inserting sentences to construct texts with the scaffolding of lexical cohesion and sentence relations. The students’ actions in text construction were recorded in a recording module which showed their metacognitive process during the task. Results of this study revealed that the engagement in metacognitive process was rarely found in the students who made little progress through three online texts. The students who demonstrated the developmental progress in three texts were actively engaged in the metacognitive process to monitor and evaluate their text constructive problems. They then frequently revised the text by reselecting the sentences with the assistance of lexical cohesion and sentence relations. It was further found that students’ metacognition could be strengthened by the text construction task. The increased engagement in metacognition, especially reselecting the sentences (revising), led to higher scores being obtained in text construction.

Keywords
Metacognition; Lexical cohesion; Online text construction; Reading process; Reading strategy

Introduction
Text construction is regarded as an important task in learning since it examines students’ awareness in text structure and coherence. The sentence-shuffling exercise, for example, is derived from the notion of the text construction and has been widely used in learning (Johns & Lixun, 1999). The shuffled sentences from a text are provided for students to establish the sentence continuation for text coherence. When students construct the meaning of the text to establish a comprehensible paragraph, they undergo the ongoing monitoring and evaluating process that might possibly result in the reformulation of cognitive and sentence structure. The process of text construction involves students’ exercising their meaning-making, connection-building, and relationships-constructing abilities to organize the selected information to construct a coherent text. In text construction, students need to integrate related units in their memory to enable comprehension of the text to be constructed (Charney, 1994; Ruiz-Primo, 2004). According to the results of Bensoussan and Laufer’s study (1984), one of the challenges that EFL students face is to identify the connections between sentences in a text. Other studies (e.g., Chu, Swaffar & Charney 2002; Sharp, 2003; Wang and Ding, 1998) also reveal that many Taiwanese EFL students fail to recognize the lexical cohesion ties, defined as a property of text, to connect sentences in integrating textual information, which results in their comprehension breakdown.

Metacognition in text construction
Metacognition is reported to benefit students in monitoring and regulating their meaning-constructing process (Azevedo, Cromley, Winers, Moos, & Greene, 2005; Block, 1992; Baker & Brown, 1984a, 1984b; Brown, 1981; Paris & Winograd, 1990; Hartman, 2001). It is further identified as the ability that modulates the cognitive process and involves "active monitoring and consequent regulation and orchestration of cognitive process to achieve cognitive goals" (Flavell, 1976, p. 252). Schraw (2001) concludes that “metacognition consists of knowledge and regulatory skills that are used to control one’s cognition” (p. 6). Li and Munby (1996) investigate two ESL readers’ cognitive process in reading academic texts. It is reported that metacognition assisted readers to resolve problems they had encountered during reading.

Brown (1987) specifically characterizes four components in metacognition: planning, monitoring, evaluating, and revising. Planning refers to the deliberate activities that organize the entire learning process, including setting the goal, sequence, strategies, and expected time for learning. Monitoring involves the activities that moderate the current progress of learning. For example, students might ask themselves questions, such as “Am I on the right...
track?”, “What should I do?”, “What information is important to complete the given tasks?”, “Should I adjust my strategies?” (Azevedo et al., 2005; Lee & Baylor, 2006).

Evaluating one’s own learning process involves an assessment of the current progress of the learning activity. This evaluation can assist students to develop the necessary skills and strategies. In the text construction, to evaluate comprehension, students have to hold the concepts acquired from the text in mind while adjusting the concepts upon subsequent information (Baker & Brown, 1984a, 1984b). When inconsistencies occur, regulation of comprehension enables students to check the text again and to take remedial actions with metacognitive strategies, such as rereading or revising, to solve inconsistencies.

Revising one’s own learning process refers to modifying previous strategies in order to achieve one’s goals. Revising is one of the regulatory actions employed to solve the problems. These four components of metacognition all lead to the improvement of comprehension (Brown, 1987; Azevedo and Cromley, 2004). Hartley (2001) indicates that students would not engage in monitoring and regulating their learning unless they are asked to do so through some activities, such as text construction.

Related studies on text construction

The task of computer-aided text construction provides students with chances to manipulate shuffled sentences and reconstruct them in a coherent way. One of the sentence-shuffling programs, Jumbler, was developed and required students to restore the original text from the scrambled words and sentences (Johns & Lixun, 1999). Two kinds of reordering were included: (1) a sentence was randomly picked and the words in it were reordered, and (2) all sentences in a text were reordered. These were further recorded in log files to show the end product of students’ reordering. From log files, teachers could identify which sentences were more difficult than the others for students to place correctly. As Johns & Lixun (1999) point out, “detailed analysis of enough log files should be able to show the advantages and disadvantages of different strategies,” and “problems were identified by studying the log file: how sentences were moved during the construction; how much time was spent for each move” (p. 335).

Higgins, Lawrie & White (1999) established the Jumbler program, SEQUITUR, with the aim of developing students’ awareness of cohesive devices and text coherence by showing students the beginning sentence of a text and randomly offering them possible continuations of text-construction. SEQUITUR also collects students’ decisions and time by log files. However, they indicate that “we do not know for sure what is happening during the thinking time,” and “readers will in most cases be able to reject the wrong continuations through an obvious lack of cohesion, but sometimes they will need to look for evidence of coherence which requires careful thought or reading ahead for several lines” (p. 342). In other words, the cognitive process that was involved in reordering sentences in a text remained unknown in Higgins et al’s system.

Many studies have reported on the results of how students constructed texts (e.g. Folkesson & Swalander, 2007; Glenberg, Brown, Levin, 2007), yet the cognitive process of how texts were constructed and how the constructive process was formulated and reformulated, have rarely been addressed. It was hard to examine students’ developmental process of meaning construction as the methods, such as naturalistic observation, interviews, or think-aloud protocols, were either too time-consuming, labor intensive (Schacter, Herl, Chung, Dennis, and O’Neil, 1999; Yeh, Yang, and Wong, 2010) or disruptive to the learning process. Many studies have sorely reported on the end results of text construction by using achievement tests or pre- and post-tests, which provided little information about what really went on in students’ minds as well as their developmental processes (e.g. Folkesson & Swalander, 2007; Glenberg, Brown, Levin, 2007; Yang, Yeh & Wong, 2008; Yeh, Yang, and Wong, 2010). The learning results in the previous studies were solely presented in the students’ end products without indicating what factors got involved in their metacognitive processes.

Purpose of the study

This study reports on a text construction process where EFL university students insert sentences to construct paragraphs in texts with the scaffolding of lexical cohesion and sentence relations. In the system, students should first understand the goal of text construction in order to select strategies in establishing the consistency between
sentences. This is the process of planning in this system as students need to know the goal of the text construction task and what they need to do next. Students then have to monitor their selections of lexical cohesion and sentence relation to approach the ultimate goal of inserting correct sentences. After monitoring, students have to evaluate four optional sentences and select a correct one related to the previous sentence. After the submission of an inserted sentence, students are encouraged to revise the incorrect sentence into the correct one. The relationship between students’ actions in the online system and corresponding metacognitive strategies is shown in Figure 1.

<table>
<thead>
<tr>
<th>Students’ actions in the system</th>
<th>Metacognitive strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understand objectives of the text construction</td>
<td>planning</td>
</tr>
<tr>
<td>Select types of lexical cohesion &amp; inter-sentential relations</td>
<td>monitoring</td>
</tr>
<tr>
<td>Select multiple choice items for inserting sentences</td>
<td>evaluating</td>
</tr>
<tr>
<td>Revise the incorrect sentence to the correct one</td>
<td>revising</td>
</tr>
</tbody>
</table>

*Figure 1. The relationship between students’ actions and metacognitive strategies*

Students’ metacognitive strategies of monitoring, evaluating, and revising were recorded in the recording module. As they were involved in the process of making sense of texts and making inter-sentential connections by lexical cohesion, their engagement of metacognitive and developmental progress could be further analyzed. The process data can be used to examine students’ actions in constructing meaning of the texts (Langer, Bartolome, Lucas, & Vasquez, 1990) where comprehension is developed and altered throughout the text construction task. Two major research questions are addressed in this study: (1) How do students with different proficiency levels construct online texts with the scaffoldings of lexical cohesion and sentence relation? (2) To what extent does metacognition facilitate students with different proficiency levels to construct online texts?

*Figure 2. The frequency distribution of the students’ scores*

**Method**

**Participants**

Two reading classes in Vocabulary and Reading were randomly selected from a university of science and technology in central Taiwan. In these two classes, the 83 college freshmen were expected to choose and practice reading strategies for enhancing their text comprehension. A nationwide screening test with the maximum score of 100 was administrated by the university to identify their reading proficiency. The frequency distribution of these 83 participants’ scores is shown in Figure 2. The score interval 71-80 provides the benchmark for dividing participants into three proficiency levels. Participants with reading scores below 71/100 were identified as less-proficient readers.
and participants whose reading scores above 80/100 were grouped as more-proficient readers. Participants whose
reading scores fell in the score interval 71-80 (19 students) were excluded from this study. Thus, 33 more-proficient
readers and 31 less-proficient readers were identified in this study. The mean score of the more-proficient readers is
92.73 with a standard deviation of 4.48. The mean score of the less-proficient readers is 66.80 with a standard
deviation of 11.67. Student I and student II were randomly selected from each less-proficient and more-proficient
reading group to illustrate how each of them made progress in text construction.

Materials

In order to provide English texts for the EFL students to construct, the study adopted the following criteria for the
selection of the texts. The three texts were similar in length in terms of the number of sentences and paragraphs.
They were all selected from the following texts: College Reading Workshop (Malarcher, 2005) entitled The Best
Travel Bargains (Text 1), with 168 words and 6 lexical cohesive pairs; Traditional Markets VS. Modern Markets
(Text 2), with 206 words and 15 lexical cohesive pairs; and Fat to Store or Fat to Burn (Text 3), with 188 words and
13 lexical cohesive pairs. The texts have similar difficulty levels of between 10.2 and 10.5 on the Flesch-Kincaid
grade level.

System Development

The system built for this study includes three modules: user interface, recording module, and feedback module. The
relationships among these three modules are presented in Figure 3. In the user interface, the teacher sets the
objectives, selects texts and enters the texts into a database. The students construct texts with the scaffolding of
lexical cohesion and sentence relation. The recording module documents the students’ constructive behavior and
process. The feedback module calls WordNet 2.1 (Dehne, Steuten and van de Riet, 2001), matches the lexical
cohesive items, and provides candidate words back to the students when they have difficulty identifying the lexical
cohesion.

User Interface

The user interface includes a teacher interface and a student interface. The teacher interface allows the teacher to
manage a course, provide the texts to be constructed, and analyze the students’ constructive process and behavior.
The student interface allows a student to undertake text construction tasks.
As shown in Figure 4, the student interface is divided into five areas. The main text area (1) presents the text that is being constructed. An “Insert” button tagged with a serial number represents a missing sentence that a student should identify based on the first and the last sentences (in boldface) of a paragraph. A “Revise” button allows the students to regulate their comprehension and revise a sentence before final submission. Next, the area of multiple-choice items (2) provides four sentences, tagged A, B, C, and D for the students to select from so that they could evaluate their comprehension. Three distracting sentences were randomly picked from the main text except for the first and the last sentences. In the area of lexical cohesion (3), the students are purposely asked to fill in two cohesive items from the current sentence and its preceding sentence to help them raise their metacognition and build sentence consistency. Once they have difficulty in selecting correct cohesive items, they can get hints of cohesive words by clicking the “Search” or “Answer” buttons that are created based on the number of cohesive word-pairs embedded in the two sentences.

In the area of cohesion types (4), the students have to fill in the relations between the two cohesive items from the text fields A and B. The lexical cohesion types include: repetition, synonym, hypernym, hyponym, meronym (Halliday and Hasan, 1976). From the menu of inter-sentential relations, the students have to choose the relation between the two sentences. The relations include: addition, clarification, comparison, contrast, example, location or spatial order, cause and effect, summary and time (Sharp, 2003).

**Recording module**

The system uses a recording module to trace the students’ constructive process and behavior that indicate their developmental progress. The teacher can analyze the comparison table between the expert’s and the student’s. The records are also helpful for the teachers to modify their instruction according to the students’ strengths and weaknesses.

**Feedback module**

WordNet 2.1 is used in the current computer system to find lexical cohesion including “repetition,” “synonyms,” “hyponym,” “hyperonym,” and “meronym.” It also assists the students to identify the relationships between the lexical cohesive items. In this study, the “SEARCH” button is designed to scaffold the students’ text construction.
The candidate lexical cohesion words generated from WordNet are shown in red color in both the main text and the multiple-choice items.

As shown in Figure 5, one of the scaffolding features provided by the system is the search button in the feedback module. For example, when a student enters the word “move” in text field A and chooses the type of lexical cohesion, the cohesive words found in WordNet will be highlighted as candidate cohesive words, like travel and move as shown above. The students can click the “ANSWER” button to get possible pair(s) of lexical cohesive items, like [United States, country] and [move, travel], when they cannot figure out the cohesive items (Figure 6).

Procedures for Data Collection

Eighty-three students were asked to construct three texts online over three different periods of time with one-month intervals between September 20th and November 26th. Before the instructors introduced the strategy of how to identify lexical cohesion in text comprehension, the students undertook the text construction task for the first time. They were asked to complete the second text after an interval of one-month and to finish the final text after two-months of instruction. The text construction tasks were to identify the target sentence from the multiple choices, pairs of lexical cohesion and their corresponding type, and finally the sentential relation. The online computer system graded the students’ text-construction by giving one score point to each correct answer of (1) sentences of the constructed text, (2) pair of cohesive words, (3) type of lexical cohesive ties, and (4) sentence relation.

Procedures for Data Analysis

The collected data were analyzed in terms of the students’ scores and process in text construction. The students’ scores refer to their correct selection of (1) sentences of the constructed text, (2) pair of cohesive words, (3) type of lexical cohesive ties, and (4) sentence relation. A text construction process includes the students’ constructive
process and behavior traced in the recording module. The trace results reveal the students’ difficulty in the constructive process and how they improved their understanding of the text. Expert maps of lexical cohesive ties were used to compare with those of the students’.

Four students’ trace results, two from each group of the less- and more-proficient students in their text construction, were randomly selected for data presentation. In the less-proficient group, student III represents how some students did not make further improvement whereas Student IV represents some students who took a further step for the process of revising and making corrections in text construction. In the more-proficient group, student V represents those who obtained the correct answers at the first attempt and who also reread the texts to undertake actions to detect the text consistency by providing extra pairs of correct lexical cohesion. Student VI represents some students who used metacognitive strategies to construct and reconstruct the online texts. An interview was also conducted with these four students individually to explore the key factors that facilitated their online text construction.

Result

Students’ scores in text construction (N=83)

The current study focused on understanding the students’ metacognitive and developmental process in constructing online texts. The data of 83 participants’ learning results in the three texts were presented in Table 1. The total score for sentence selection in Text 1 was 4 and the score for lexical cohesive ties was 6. For example, 1.36/4 in the table represents the mean score for all students (1.36) divided by the full score 4, which equals 34.04%. The calculation is the same for Text 2 and Text 3.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Text 1</th>
<th>Text 2</th>
<th>Text 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Css* mean</td>
<td>Clt** mean</td>
<td>Css* mean</td>
</tr>
<tr>
<td>All students</td>
<td>1.36/4 (34.04%)</td>
<td>1.16/6 (19.28%)</td>
<td>3.80/7 (54.22%)</td>
</tr>
<tr>
<td>More-proficient</td>
<td>2.12/4 (53%)</td>
<td>1.64/6 (27.33%)</td>
<td>5.39/7 (77%)</td>
</tr>
<tr>
<td>students</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less-proficient</td>
<td>0.77/4 (19.25%)</td>
<td>0.87/6 (14.5%)</td>
<td>2.77/7 (39.57%)</td>
</tr>
</tbody>
</table>

As shown in Table 1, most students made progress in three text construction tasks as the percentage of correct sentence selection increased from 34.04% to 55.02%. For the less-proficient students, the percentage of correct sentence selection increased from 19.25% to 43.5%. This was also true for their performance in lexical cohesive ties. The less-proficient students progressed from 14.5% (Text 1) to 37% (Text 3).

Developmental Process of the students’ text construction

The students’ process data derived from the trace result was analyzed to look into the students’ developmental process through constructing different texts. A mental map was drawn based on what lexical cohesive ties the students had identified in the system to compare against the expert map. Table 2 shows an example of Student I’s text construction with lexical cohesion in Text 1. The following is Text 1 with a serial number inserted into each sentence. The underlined words are lexical cohesive ties.

Sample Text:

The United States has a reputation as a society on the move. (2) Not only do people and families travel across the country for new jobs or educational opportunities, but they also take yearly vacations. (3) Setting up the plans for a trip can be complicated, though. In order to get the best price on airline tickets, people have to be willing to put in a little effort.
One reason that finding good prices for travel is so complicated is because airlines have complex formulas for inventory management so they can maximize profits by filling planes. When there are lots of reservations (during peak seasons), these companies can charge higher prices and still be sure that somebody will need their services no matter how much it costs. On the other hand, during the off-peak season, demand is low, so companies cut their prices to try and attract people who would normally not travel at that time. One good place in which to find these last-minute bargains is on the Internet.

As shown in Table 2, the expert map represented the correct pair of lexical cohesion whereas the student’s map is shown on the right column. Blank slots were used to indicate the incorrect answers identified by the student. Each pair of cohesive words in two connected sentences (e.g., sentence 1 and 2) was circled and presented as the sentence number, such as (1, 2). In other words, the expert map in Table 2 showed that there were 2 pairs of lexical cohesion [United States, country] and [move, country] between sentences 1 and 2.

For each inserted pair of lexical cohesion, Student I missed all of the seven pairs of the cohesive words in Text 1. The correct rate for lexical cohesion is 0%. For Text 2, only two correct pairs of lexical cohesion, mainly repetition, were identified by student I (Table 3). It was found that he achieved little understanding of lexical cohesion. His correct rate in lexical cohesive ties of Text 2 was 12%. Finally, in Text 3, student I still remained unaware of the lexical cohesion and he could identify only one pair of correct cohesive words. His correct rate in lexical cohesive ties of Text 3 was 7%. He did not use lexical cohesion to enhance his comprehension of the text. As he had limited understanding of lexical cohesion, the correct pairs of cohesive words chosen were mainly “repetition.”

### Table 2. The comparison between the expert’s and Student I’s maps in Text 1

<table>
<thead>
<tr>
<th>The expert’s map</th>
<th>Student I’s map</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Expert Map" /></td>
<td><img src="image" alt="Student Map" /></td>
</tr>
</tbody>
</table>

### Table 3. The comparison between the expert’s and Student I’s maps in Text 2

<table>
<thead>
<tr>
<th>The expert’s map</th>
<th>Student I’s map</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Expert Map" /></td>
<td><img src="image" alt="Student Map" /></td>
</tr>
</tbody>
</table>
The results of Student II differed from Student I who could only identify repeated words; Student II demonstrated a better development in incorporating lexical cohesion. Based on the graphic representations, it was found that she identified only one pair of lexical cohesion correctly (17%) in Text 1 (Table 5). Gradually, she raised the awareness of lexical cohesion in Text 2 (Table 6). As shown in Table 6, Student II made progress in her text construction. Although some pairs of lexical cohesion were still not identified, Student II increased her understanding in Text 2 and found eight correct pairs of lexical cohesion (47%). Her correct rate of lexical cohesion pairs increased from 17% in Text 1 to 47% in Text 2. Student II continued to make progress in their text construction in Text 3 (Table 7), by missing only one pair of lexical cohesion between sentence 1 and 2. In Text 3, her rate in identifying lexical cohesion correctly was 92%. The comparison tables in the three texts indicated that Student II made gradual progress in understanding lexical cohesion.
Table 7. The comparison table of the expert’s map and Student II’s map in Text 3

<table>
<thead>
<tr>
<th>Text 3: Fat to store or fat to burn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The expert’s map</strong></td>
</tr>
<tr>
<td><img src="image1.png" alt="Expert's Map" /></td>
</tr>
</tbody>
</table>

The students’ metacognitive process in text construction

Two students in each group of the less- and more proficient students were randomly chosen to illustrate how some students make progress and some do not. In the less-proficient group, student III represents those who did not make progress and student IV represents those who did. In the more-proficient group, student V represents those proficient readers who did not make further progress since she obtained the correct sentence selections at her first attempt. She then read and reread the text by providing one more pair of lexical cohesion. Student VI represents those who made progress by revising the incorrect sentence selections to the correct ones by making good use of scaffoldings.

The example below presents the comparison between Student III and Student IV (from the less proficient group) and Student V and Student VI (from the more proficient group) in constructing one of the texts (see the sample text on page 102). The numbers in Figures 7a, 7b, 7c, and 7d refer to the sequence of actions that Student III undertook in text construction. In action [1] of Figure 7a, Student III filled in Brazil and Country, as synonyms, and addition as the inter-sentential relation. In this study, this action corresponds to monitoring in the process of metacognition as Student III did not achieve the ultimate goal of inserting a correct sentence. Student III merely monitored whether he was on the right track to approach the ultimate goal. In action [2], Student III selected one sentence from the four given choices to insert in the text. It corresponds to evaluating in the process of metacognition as Student III had to evaluate these four choices and select a correct one.

![Figure 7a. Student III’s actions in inserting sentence #2](image3.png)

![Figure 7b. Student III’s actions in inserting sentence #3](image4.png)
Based on the actions that Student III took from Figure 7a to Figure 7d, his metacognitive process in text construction is shown in Figure 8; the numbers refer to the sequential actions that Student III went through in the process of metacognition. It was found that Student III repeated the process of monitoring and evaluating and did not take a further step for the process of revising.

**Figure 8.** Student III’s metacognitive process in inserting sentences

Student IV undertook the monitoring and evaluating process to re-select the sentence relation types in actions [3] and [7] and revise the incorrect inserted sentences to the correct ones in actions [4] and [8].

**Figure 9a.** Student IV’s actions in inserting sentence #2
According to the actions that Student IV took from Figure 9a to Figure 9d, his metacognitive process in text construction is presented in Figure 10. Student IV monitored a correct sentence relation (action [3]) and took a further step to revise (action [4]). As indicated in Figure 10, the curved arrow 4 goes from monitoring to revising. Student IV continued to complete lexical cohesion and sentence relation (action [5]) for sentence #3 in which the
curved arrow 5 goes from revising to monitoring. It was found that Student IV not only underwent the monitoring and evaluating process (action [2] [3] [6] & [7]) to identify lexical cohesion and sentence relation but also went through the revising process (action [4] and [8]) to correct sentence relations and further revise the incorrect sentences into the correct ones. In sentence #4 and #5, student IV did not make any revision.

Student V had obtained the correct sentence selections so she did not undertake any revision in her sentence selections. She repeated the monitoring and evaluating process to re-read the text and provided one more pair of correct lexical cohesion. It demonstrates that she tried to reread the text consistency and detected the lexical cohesion between sentences.

<table>
<thead>
<tr>
<th>[1] Lexical cohesion</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>Brazil</td>
<td>country</td>
<td>hyponym</td>
</tr>
<tr>
<td>[1] Sentence relation</td>
<td>Addition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2] Insert sentence #2</td>
<td>In these cities, street markets, or feiras, are where seventy percent of the fruits and vegetables bought in the country are sold</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 11a. Student V’s actions in inserting sentence #2

<table>
<thead>
<tr>
<th>[3] Lexical cohesion</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>Brazil</td>
<td>country</td>
<td>hyponym</td>
</tr>
<tr>
<td>citys</td>
<td>citys</td>
<td>repetition</td>
<td></td>
</tr>
<tr>
<td>[1] Sentence relation</td>
<td>Addition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2] Insert sentence #2</td>
<td>In these cities, street markets, or feiras, are where seventy percent of the fruits and vegetables bought in the country are sold</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 11b. Student V’s actions in revising sentence #2

Figure 12. Student V’s metacognitive process in text construction

Student V’s metacognitive process in text construction is shown in Figure 12 according to the actions she took from Figure 11a to Figure 11b. It was found that Student V moved back and forth between the monitoring and evaluating process, and added one more pair of lexical cohesion. In inserting sentence #3, #4 and #5, she also got the correct sentence selections and further provided one more pair of lexical cohesion like the actions she took in sentence #2. In other words, Student V represented those students who got the right answers in first place and also reread the texts to provide extra pairs of correct lexical cohesion.
Student VI represented students who actively used metacognitive strategies to construct and reconstruct the online texts. The example below presents how she selected types of cohesion, inter-sentential relations, and multiple choice items for inserting sentences (from Figure 13a to Figure 13g). She also asked for candidate words from WordNet to revise the sentences in Figures 13b, 13e and Figure 13g.

<table>
<thead>
<tr>
<th>[1] Lexical cohesion</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>Brazil</td>
<td>country</td>
<td>hyponym</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[1] Sentence relation</th>
<th>correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[2] Insert sentence #2</th>
<th>When an international chain store moves into a new country, it should keep this lesson in mind</th>
</tr>
</thead>
<tbody>
<tr>
<td>incorrect</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13a. Student VI’s actions in inserting sentence #2

<table>
<thead>
<tr>
<th>[3] Lexical cohesion</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>Brazil</td>
<td>country</td>
<td>hyponym</td>
</tr>
<tr>
<td>cities</td>
<td>cities</td>
<td>repetition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[3] Sentence relation</th>
<th>correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[4] Revise sentence #2</th>
<th>In these cities, street markets, or feiras, are where seventy percent of the fruits and vegetables bought in the country are sold</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13b. Student VI’s actions in revising sentence #2

<table>
<thead>
<tr>
<th>[5] Lexical cohesion</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>incorrect</td>
<td>markets</td>
<td>shopping</td>
<td>Hypoynym</td>
</tr>
<tr>
<td>markets</td>
<td>supermarket</td>
<td>hypernym</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[5] Sentence relation</th>
<th>correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[6] Insert sentence #3</th>
<th>These shopping habits have led people to shop at the supermarket once a month and at the feira once a week</th>
</tr>
</thead>
<tbody>
<tr>
<td>incorrect</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13c. Student VI’s actions in inserting sentence #3

<table>
<thead>
<tr>
<th>[7] Lexical cohesion</th>
<th>Word 1</th>
<th>Word 2</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>feira</td>
<td>feira</td>
<td>Repetition</td>
</tr>
<tr>
<td>vegetables</td>
<td>vegetables</td>
<td>Repetition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[7] Sentence relation</th>
<th>correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[8] Insert sentence #4</th>
<th>While shopping at the feira, people can pick up the things they might want to buy, smell the fruit and vegetables, and sometimes even taste them before they buy anything</th>
</tr>
</thead>
<tbody>
<tr>
<td>incorrect</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13d. Student VI’s actions in inserting sentence #4
Student VI took remedial actions to frequently revise the incorrect lexical cohesions, sentence relations, and sentences in actions 4, 10, and 14. In revising incorrect sentences, she searched candidate words from WordNet (graph (b) and (g) of Figure 13). This also indicates how Student VI actively asked for feedback to correct her misconceptions in the process of revising.
These revisions resulted in her later success in inserting correct sentences. Based on the actions that Student VI took from Figure 13a to Figure 13g, her metacognitive process is shown in Figure 14. For example, Student VI’s action [10] was to revise an inserted sentence (action [8]) after monitoring a correct lexical cohesion and sentence relation (action [9]); the curved arrow 10 goes from monitoring to revising in Figure 14. After revising (action [10]), Student VI continued to complete lexical cohesion and sentence relation (action [11]); the curved arrow 11 goes from revising to monitoring. Comparing Figure 8 and Figure 14, it was found that the more actively a student was engaged in metacognitive process, the more correct the sentences that she selected were.

The relationship between scaffolding and the students' learning results

The frequency of the students’ request of lexical cohesive items from WordNet was automatically calculated by the Recording Module in this system. Student III asked for help from WordNet only two times, whereas Student VI requested help 14 times (Table 8). Student VI not only asked for more scaffoldings but also obtained higher scores than Student III in the three texts (Table 9). That is, Student VI underwent an active process of constructing and reconstructing the text by seeking assistance from the system, taking actions to revise the lexical cohesive ties, sentence relation, and the correct sentences, and making an effort to figure out the role that lexical cohesion plays to enhance her comprehension of the text to be constructed.

| Table 8. The frequency of Student III and Student VI requested by scaffolding in the three texts |
|---------------------------------------------------------------|---------------|---------------|
| Frequency of using WordNet                                   | Student III   | Student VI    |
| Frequency of asking for correct answers in lexical cohesion   | 2             | 14            |
| Frequency of using scaffolding                               | 3             | 38            |

| Table 9. Performance of Student III and Student VI in three texts |
|---------------------------------------------------------------|---------------|---------------|
| Text #                                                        | Text 1 | Text 2 | Text 3 | Text 1 | Text 2 | Text 3 |
| correct sentence selection                                     | 1      | 1      | 1      | 2      | 4      | 5      |
| lexical cohesion selection                                     | 0      | 2      | 1      | 1      | 8      | 12     |
| sentence relation                                              | 2      | 1      | 2      | 2      | 4      | 6      |
| Total Score                                                   | 3      | 4      | 4      | 6      | 16     | 23     |

The key factors that facilitate the students' text construction

From the interviews, student VI identified some key elements facilitating his comprehension. First, she expressed that she obtained a deeper understanding through the instruction to identify the lexical cohesion across sentences and to determine the relationship between two sentences. She was also supported with immediate and individualized feedback from the system. In the system, after a student submitted the text that had been constructed, the system compared the student’s answers with the correct ones and indicated the mistakes with the mark X in the right column (Figure 15). In the left column, it provided the student with both the correct sequence of the sentence and correct pairs of lexical cohesion as immediate feedback. With the individualized feedback, student VI could identify her own errors in selecting the lexical cohesion, the sentence relation, and the inserting sentence. When there was comprehension breakdown, she would reread the previous sentences, used the tools in the system, such as WordNet, to verify her understanding, and undertook the revision when necessary.

Student III in the less-proficient group stated that he were overwhelmed with both the immediate feedback and WordNet. He claimed that even with the provided assistance from WordNet, many candidate words were not recognizable, and therefore he did not ask feedback from the system. Student IV stated that since he was required to select sentence relations, he would try to reread the sentences many times to detect text consistency. He also claimed that the WordNet did not help him much in selecting the lexical cohesion because he could not understand some of the candidate words. Student V in the more-proficient group indicated that she would be able to detect more pairs of lexical cohesion after she read and reread the text.
Conclusion

The design of the online text construction system in this study intends to foster students’ metacognition. In general, all of the students made progress in the three text construction tasks as the percentage of correct sentence selection increased from 34.04% to 55.02%. The improvement for their performance in lexical cohesive ties was also evident as the percentage of correct lexical cohesive ties increased from 19.28% to 38.18%. Expert maps were used to compare against two exemplary students’ developmental and metacognitive process (student I & II) in constructing three online texts.

The students’ text construction process enhanced by their metacognition was shown and reported, suggesting that the students’ engagement of metacognition could be further strengthened. The findings also confirm the studies of Romainville (1994) and Taraban, Rynearson, and Kerr (2000), which indicate that metacognition is positively associated with college students’ comprehension. In this study, it was found that Student VI (the case from the more-proficient) frequently asked for candidate words from WordNet (14 times). With the scaffold of WordNet, she further selected the correct pairs of cohesive words so as to help replace incorrect sentences with the correct ones. As Student VI actively exercised her metacognition to establish the consistencies between sentences, she made good use of the scaffolding tool provided by the system to search for cohesive ties. As a result, she gradually improved her own comprehension by rereading the sentences, building the sentence relationship, and taking the revising actions. She was engaged in the metacognitive process of planning, monitoring, evaluating, and revising.

Generally, most students made progress with the strategy instruction incorporating the online system to read and reread the text (i.e. the example of student IV, V, VI). Each sample case shown in this study was to illustrate how some of the students who made progress and those who did not. It was found that students’ language proficiency and metacognitive level were the key factors that resulted in the differences in students’ text construction. The finding corresponds to the existing literature in reading strategy indicating that more-proficient students are able to use metacognitive strategy, such as monitoring and revising effectively and appropriately in a context (Anderson, 1991; Paris and Meyer, 1981; Waxman and Padron, 1987).

In this study, some students used the objectives of the system to plan for the learning goals and to understand the strategy employment, lexical cohesion and sentence relations (planning). In the process of achieving their ultimate goals, they selected cohesive words, types of cohesion, and inter-sentential relations to monitor their understanding of the text coherence (monitoring). They also had to evaluate four choices of sentences and select a correct one in the
process of evaluation (evaluating). This process differs from that of monitoring for correct sentence insertion, which was the students ultimate and only goal in text construction. After the submission of inserted sentences, the students were allowed to revise their sentences (revising). From the results of this study, it was found that the more students were engaged in metacognition, the higher they scored in text construction. In the process of revising, students had to decide whether they would search for candidate words from WordNet, which is designed as one of the scaffoldings to arouse students’ metacognition. The process of revision could be stated as self-scaffolding. As Holton and Clarke (2006) characterised self-scaffolding as referring to “situations in which an individual is able to provide scaffolding for himself when any problem or concept that is new to the individual is being tackled” (p. 136). It is an essential tool for students to solve significant problems independently.

Self-scaffolding must be developed for each student’s learning style, level, and interest if independent learning is to be achieved. A teacher can be a facilitator and provider of learning strategies to nurture students to become independent learners. A teacher should constantly monitor students’ progress and contribute in a number of ways in assisting students to exert their metacognitive strategies. Masters and Yelland (2002) stated that “the teacher must also know when to withdraw support in order to allow students to explore and construct new understandings” (p. 321). This implies that the trace results presenting students’ actions and thinking process in the online system are useful for the teachers to adjust their teaching to adapt to students’ needs. Learning should be seen as a long-term development effort that seeks to improve students’ knowledge for new requirements in later academic performance.

Some limitations were also found in this study. First, although the recording module documented students’ actions in inserting, revising sentences, searching for WordNet, the reasons for arousing their actions in revision were not fully understood. This could make the process data incomplete in this study. Second, some less-proficient students could not clearly identify the reasons why they could not make progress in text construction. It might take longer for the less-proficient students to grasp the main ideas and exert their metacognition. Finally, the process of text construction could be an extremely complicated phenomenon. This study only highlights how students use some key elements, such as lexical cohesion and sentence relation to construct the text. Further studies could possibly look into other elements in text construction.

Acknowledgment

This study was supported by National Science Council in Taiwan (NSC 97-2410-H-224-017-MY2). The research grant made the continuation of this study possible. Our gratitude goes to the anonymous reviewers for their useful suggestions.

References


A Systematic Approach for Learner Group Composition Utilizing U-Learning Portfolio

Yueh-Min Huang1,2* and Ting-Ting Wu1

1Department of Engineering Science, National Cheng Kung University, Taiwan // 2Department of Applied Geoinformatics, Chia Nan University of Pharmacy and Science, Taiwan // huang@mail.ncku.edu.tw // n9897104@mail.ncku.edu.tw

*Corresponding Author

ABSTRACT

A context-aware ubiquitous learning environment allows applications to acquire diverse learning behaviors of u-learners. These behaviors may usefully enhance learner characteristics analysis which can be utilized to distinguish group learners for further instruction strategy design. It needs a systematical method to analyze u-learner behaviors and utilize learner characteristics for group composition. This paper proposes an effective and systematic learner grouping scheme containing transformation processes from u-portfolios to the proposed Portfolio Grid, creating a learner similarity matrix, and group composition. This study also evaluates intra-group diversity of each resultant heterogeneous group and analyzes learning behavioral patterns acquired from the study experiment. The results indicate that the proposed learner grouping algorithms had positive effects on group composition and interaction between group members for follow-up ubiquitous collaborative learning.

Keywords
Ubiquitous learning, collaborative learning, group composition, learning portfolio

Introduction

During the past several years, educational research has increasingly focused on collaborative learning (CL) pedagogies (Johnson & Johnson, 1999). Many Researches have demonstrated collaborative learning as an effective teaching approach that encourages student learning with high-level cognitive strategies, critical thinking, and positive attitudes (Hsu, 2003; Johnson & Johnson, 1998), benefitting students in terms of achievement, motivation, and social skills (Johnson & Johnson, 1989; Huang, Huang & Fu, 2011).

Collaborative learning integrating computer-based information technology has transformed the learning environment into Computer Supported Collaborative Learning (CSCL) (Inkenet et al., 1999). Rapid development progress of wireless communication and continuing growth of mobile handheld devices has led learning into Mobile Computer Supported Collaborative Learning (MCSCCL) (Danesh et al., 2001), even Ubiquitous Computer Supported Collaborative Learning (UCSCL) (Hwang, Hsu & Huang, 2007) using the concepts of a novel Ubiquitous Learning (U-learning) environment (Ogata & Yano, 2004). In such a learning environment, it is possible to actively provide ways for identifying right collaborators, right contents and right services in the right place at the right time according to the individual surrounding context information of learners (El-Bishouty, Ogata, & Yano, 2007).

Several studies have indicated that learner group composition has become a fundamental issue in collaborative learning. Many researches (Hooper & Hannafin, 1988; Lin, Huang & Cheng, 2010; Webb, 1982) investigating this subject have shown that different grouping criteria for small groups affects learning performance and social behaviors of grouped members. Heterogeneous group composition not only enhances elaborative thinking, but also leads learners to deeper understanding, better reasoning abilities, and accuracy in long-term retention (Johnson and Johnson, 1999). Webb & Palincsar (1996) further suggested that group composition formed with regard to heterogeneity of members’ gender, ability, achievement, social economic status, or race, facilitates heterogeneous group composition in collaborative learning. Since a considerable number of researchers have suggested that heterogeneous grouping promotes positive interdependence, better group performance and effective interaction, this study proposes a heterogeneous grouping method to apply in a ubiquitous collaborative learning environment.

In the current information age, many learning systems provide useful functions for gathering more detailed information about learner situations in learning activities helping teachers obtain a richer understanding of learner behavior (Sakamura & Koshizuka, 2005). The purposeful collection of learning records is called a portfolio, which provides evidence of a learner’s knowledge, skills, characteristics and dispositions (Sherry & Bartlett, 2005). Moreover, the learning portfolio supports learning by including an evaluation of collected evidence and reflective
commentaries of prospective learning activities (Struyven, Dochy & Janssens, 2005). Given the rapid development of technology, e-portfolios have been increasingly used as an alternative assessment tool which allows teachers, learners, and parents to understand and evaluate the learning process as well as aiding further learning and growth (Bataineh et al., 2007). Wang and Turner (2006) proposed that reflection helps students and teachers move beyond seeing the e-portfolio as a mere alternative assessment tool to appreciating its value as a learning strategy. In recent years, the e-portfolio has been used as a diagnostic tool for evaluating and reflecting students’ learning during the whole learning process (Adams, Swicegood & Lynch, 2004).

Currently, many ubiquitous learning systems provide functions to collect and record learner behaviors or learning events in learning activities. However, existing grouping methods used in outdoor ubiquitous collaborative learning environments that often utilize gender (Savicki, Kelley & Lingenfelter, 1996), ability (Saleh, Lazonder & De Jong, 2005), individual psychological features (Tian et al., 2008; Wang, Lin & Sun, 2007), or ethnicity (Cordero, DiTomaso & Farris, 1996) to form learning groups, often ignore learners’ various learning behaviors. Therefore, this paper proposes a learner grouping algorithm which utilizes the u-learning portfolio (u-portfolio). Since the u-portfolio contains a variety of learner behavior information, the proposed algorithm obtains a more appropriate learner grouping result that more precisely reflects characteristics of each individual learner. The proposed approach utilizes the concept and technique of Repertory Grid to systematically transform an original u-portfolio into a Portfolio Grid. After constructing a portfolio grid, this work builds a learner similarity matrix. Finally, this work uses the FOCUS concept for the learner grouping algorithm to rapidly distinguish learners into appropriate heterogeneous groups. Instructors can utilize outcomes for further ubiquitous collaborative learning and team working with heterogeneous participant groups.

U-Plant Learning System

System Architecture and Components

Our previously built u-learning system, U-Plant, collected and recorded the u-portfolio data (Wu, Yang, Hwang & Chu, 2008). Figure 1 shows the U-plant system architecture. The learner interface module provides a friendly and flexible interface for learner operation with mobile devices. The learning processes and learning behaviors are particularly recorded into the u-portfolio database through the learner interface module. The current study also retrieves learning contents and materials from the course database and presents it to learners through the learner interface module. The learning management module enables authorized users, such as the teacher, to configure
learning strategies and create student accounts. Learning strategy configurations such as procedures, rules, and settings are stored in the learning strategy database while the data of created student accounts are stored in the learner profile database. Both learning strategies and student account modifications using the learning management module are updated to the databases. Moreover, the inference engine is responsible for obtaining the required information from databases and making decisions about what learning materials should be retrieved from the course database and delivered to the individual learner. This decision is made by the analysis process in the inference engine with various parameters from the learning strategy, learner profile, and the u-portfolio database. The course management module not only helps the teacher manage learning content, but is also responsible for retrieving adaptive learning content to deliver to students for ubiquitous learning. Context-awareness functionality is a fundamental requirement in the U-Plant learning environment. The learning objects are equipped with tiny RFID tags. Each student takes a mobile device equipped with a passive RFID reader which detects users’ location using RFID technology.

Figure 2 shows a learning activity that guided the students to observe the leaf shape of the plant. Learners in the empirical environment of the botanical garden learned, step-by-step, the facts and special characteristics of the empirical objects of learning (plants), according to the learning content and observational steps provided by the system. Learners also used the functions of the system to have interactive discussion and Q&A sessions, record notes, and engage in other learning activities.

U-Portfolio

This study utilizes the web ontology language (OWL) technique for describing learners' u-portfolio and developing a connection between learners and services (Figure 3).

The u-portfolio ontology contains essential learner profiles (Yang, 2006) and the proposed location profile and behavior profile in U-Plant system. Location profile records location movements during students' learning activities utilizing RFID equipments. Behavior profile records u-learning behaviors classified into eight pre-defined categories:

1. Moving: the learner finishes and leaves a certain learning object, and moves on to the next learning object based on the map directions indicated in the mobile device.
2. Losing: the learner cannot successfully locate the appointed learning object to conduct learning.
3. Observing: the learner has moved to a certain learning object and started to use the learning functions in a mobile device until learning is finished and the learner exits.
4. Referencing: the learner looks at the expositions, hints, and notes shown on the learning device.
5. Answering: the learner answers a quiz about observed objects.
6. Interacting: the learner communicates with other classmates.
Completing: the learner completes a certain percentage of the learning objects in a specific amount of time.

Taking note: the learner utilizes the system support function to record some information during the learning process.

Learner ontology = \{Profiles, Environment, Devices\}

Profiles = \{Personal, Calendar, Social, Location, Behavior\}

- Personal_profile = \{student ID, name, gender, phone, address, email, role\}
- Social_profile = \{owner, collaborator\}
  - owner = \{student ID, name\}
  - collaborator = \{partner, interactive_type\}
    - partner = \{student ID, name, contact_info\}
    - interactive_type = \{one on one | working_team | community\}
- Location_profile = \{RFID_ID, site_name, arrival (yyyy:mm:dd;hh:mm), leave(yyyy:mm:dd;hh:mm)\}
- Behavior_profile = \{moving, losing, referencing, observing, answering, interacting, completing, taking note\}
  - moving = \{from, reach, time\}
    - from = \{RFID_ID, site_name\}
    - reach = \{RFID_ID, site_name\}
    - time = \{begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - losing = \{reach, miss, time\}
    - reach = \{RFID_ID, site_name\}
    - miss = \{RFID_ID, site_name\}
    - time = \{begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - referencing = \{object, reference_type, time\}
    - object = \{RFID_ID, site_name\}
    - reference_type = \{exposition | hint \ note\}
    - time = \{begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - observing = \{RFID_ID, begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - answering = \{RFID_ID, correct, begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - interacting = \{object, attendee, time\}
    - object = \{RFID_ID, site_name\}
    - attendee = \{student ID, name\}
    - time = \{begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - completing = \{ratio, time\}
    - time = \{begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}
  - taking note = \{object, behavior, content, time\}
    - object = \{RFID_ID, site_name\}
    - behavior = \{moving, losing, referencing, observing, answering, interacting\}
    - content = \{title, description\}
    - time = \{begin(yyyy:mm:dd;hh:mm), end(yyyy:mm:dd;hh:mm)\}

Environment = \{School botanical garden, Pool, Library, English classroom, Specialized classrooms\}

Devices = \{Tool, Equipment\}

- Tool = \{PDA, mobile phones, Eee PC\}
- Equipment = \{platform, CPU speed, memory size, screen size\}

Figure 3. Definition of u-portfolio ontology

Proposed Research Methodology

Portfolio Transformation

Given the flowchart presented in Figure 4, the initial stage retrieves portfolio database records that can be used as factors for distinguishing learners’ characteristics. The next step transforms the portfolio data retrieved from LMS into the Portfolio Grid, which column fields consist of a set of Elements. Table 1 shows a case of portfolio grid. Each student is listed as an Element and put in the top of the grid as a column caption.
Table 1. Example of a portfolio grid

<table>
<thead>
<tr>
<th>Trait Attributes</th>
<th>Elements</th>
<th>Ann</th>
<th>Tom</th>
<th>Jon</th>
<th>Eva</th>
<th>Joy</th>
<th>May</th>
<th>Elements</th>
<th>Opposite Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>observing for a long time</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>observing for a short time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>answering quiz correctly</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>answer quiz incorrectly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>interacting frequently</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>interacting seldom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>moving for a long time</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>moving for a short time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>losing frequently</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>losing seldom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>answering for a long time</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>answering for a short time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>referencing for a long time</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>referencing for a short time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>referencing hint frequently</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>referencing hint seldom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>referencing note frequently</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>referencing note seldom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>high degree of completing</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>low degree of completing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>taking note frequently</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>taking note seldom</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The rows of a portfolio grid consist of a set of Constructs and each Construct contains a pair of “trait attributes” and “opposite attributes”. The current work puts trait attributes on the left-hand side of the grid, and opposite ones on the right-hand side. After determining the Elements in columns and the Constructs in rows, the subsequent procedure provides suitable evaluation values and fills them into the grid cells. The evaluation values are presented using the K-scale rating mechanism which converts original data values into K-scale rating values using the following calculation formula shown in (1).

\[
r = \begin{cases} 
    K, & \text{if } v = V_{\text{min}} \\
    K - \left[ \frac{v - V_{\text{min}}}{(V_{\text{max}} - V_{\text{min}} + 1)/K} \right] + 1, & \text{otherwise}
\end{cases}
\]

(1)

where notation \( r \) is the obtained K-scale rating value by calculation; \( v \) is the retrieved average value from a database table field; \( V_{\text{min}} \) and \( V_{\text{max}} \) are the minimum and maximum average values respectively in the corresponding database table field; and \( K \) is the adopted rating scale for the rating calculation. These calculation result values that have lower
ratings indicate a significant characteristic towards the trait attribute. Contrarily, the higher one means a characteristic towards the opposite attribute. In other cases where the retrieved original data is a Boolean-type value, the TRUE and FALSE values are represented using rating values 1 and \( K \) respectively.

Create Weighted Similarity Matrix

The subsequent procedure judges the weight of each Construct listed in the grid. Different weights mean that each Construct can be assigned to a different degree of importance for similarity comparison and heterogeneous grouping. Adjustments to the setting of the weight could be made according to the teachers’ experience, or based on the difference between the design of learning activities and group strategy. If the teacher has not set any weights, the system will automatically assign the same weighted value to each construct. Additionally, the settings of the system provide an interface that allows teachers to see a preview of the results of learner groups, which is derived from various weight settings that have been processed by the system’s methods of analysis. This provides reference for the teachers in making decisions about setting the weighted value.

Table 2. Example of a similarity matrix

<table>
<thead>
<tr>
<th></th>
<th>Ann</th>
<th>Tom</th>
<th>Jon</th>
<th>Eva</th>
<th>Joy</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>100</td>
<td>30</td>
<td>50</td>
<td>75</td>
<td>66</td>
<td>36</td>
</tr>
<tr>
<td>Tom</td>
<td>30</td>
<td>100</td>
<td>57</td>
<td>18</td>
<td>23</td>
<td>70</td>
</tr>
<tr>
<td>Jon</td>
<td>50</td>
<td>57</td>
<td>100</td>
<td>39</td>
<td>57</td>
<td>68</td>
</tr>
<tr>
<td>Eva</td>
<td>75</td>
<td>18</td>
<td>39</td>
<td>100</td>
<td>59</td>
<td>25</td>
</tr>
<tr>
<td>Joy</td>
<td>66</td>
<td>23</td>
<td>57</td>
<td>59</td>
<td>100</td>
<td>52</td>
</tr>
<tr>
<td>May</td>
<td>36</td>
<td>70</td>
<td>68</td>
<td>25</td>
<td>52</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2 shows the matrix of learner similarity generated. Calculation of the weighted value for each element of the similarity matrix in this paper is shown in (2). The notation \( S_{ij} \) denotes the percentage of similarity between learners \( L_i \) and \( L_j \). The notations \( n \) and \( m \) are the numbers of Elements (learners) and Constructs in the portfolio grid. \( PG(L_i, C_h) \) denotes the cell value within column (Element) \( L_i \) and row (Construct) \( C_h \) in the portfolio grid; \( \alpha_h \) denotes the weight assigned to the Construct \( C_h \); and \( K \) is the rating scale used in the previous developmental stage of the grid.

\[
S_{ij} = 1 - \left( \sum_{h=1}^{m} \alpha_h \frac{PG(L_i, C_h) - PG(L_j, C_h)}{K - 1} \right) \times 100 \%
\]

Grouping Algorithm

After the weighted similarity matrix of learners is generated, a teacher could determine a threshold of difference or a number of groups for the use of two heterogeneous clustering algorithms in this paper.

Figure 5 shows a heterogeneous grouping with a given difference threshold \( T \). This grouping algorithm not only allows students to cooperate in a more effective manner to complete the assignment, but also enables individual group members to bring distinctive thinking to the design task. The steps of the heterogeneous grouping algorithm are below.

Step 1. Assign the difference threshold \( T \) according to the empirical rule or the learning strategy design.
Step 2. Due to heterogeneous grouping, first establish a triangular matrix \( M' \) as the difference matrix. The element value is 100 minus the element value of the similarity matrix.
Step 3. Assign every learner to be an independent group by default.
Step 4. Select the element with the maximum value (the greatest difference value) from all of the elements in the matrix \( M' \).
Step 5. If the maximum value is greater than the threshold \( T \), find the two learners (row and column of matrix \( M' \)) corresponding to the element value.
Step 6. If each of the found learners belongs to a different group, merge them into one group.

Step 7. Delete the maximum value element just selected. Repeat Steps 4-6 until a difference value greater than $T$ cannot be found in the matrix $M'$.

Step 8. Based on the obtained heterogeneous grouping results, the number of people in each group may differ. Therefore, teachers can decide whether or not to set the difference threshold $n_T$ of the number of people. If the difference between any two groups of people is greater than $n_T$, then the teacher can execute the `balance()` function to conduct minor adjustments to balance the number of people.

Figure 5. Heterogeneous grouping algorithm with given difference threshold

Figure 6. Heterogeneous grouping algorithm with given number of groups

Figure 6 shows a heterogeneous grouping with a given number of groups $G$. The steps are as follows:

given a similarity matrix $M = \{ m_{ij} \mid 1 \leq i, j \leq n \}$ where $n$ is the number of students.
given a difference threshold $T$ for the heterogeneous grouping
let triangular matrix $M' = \{ 100 - m_{ij} \mid m_{ij} \in M \text{ and } i < j \}$
let $\{s_1, s_2, s_3, \ldots, s_n\}$ denote the set of all students
$\forall i \in [1, n]$, let $C_i = \{ s_i \}$
let $C = \{ C_i \mid i \in [1, n] \}$
while TRUE
   select an element $m_{ij} \in M'$ such that $m_{ij}$ is largest in $M'$
   if $m_{ij} \geq T$
      find $C_h$ and $C_k$ in $C$ such that $s_i \in C_h$, $s_j \in C_k$
      $C_h \leftarrow C_h \cup C_k$
      remove element $C_k$ from $C$
      set $m_{ij} = \text{NULL}$
   else
      exit while
   end if
end while
if $\exists C_i, C_j \in C$ and $|C_i| \cdot |C_j| > n_T$ // $n_T$ indicates the group size error threshold
   call `balance(C, T, n_T)`
end if
// $C$ represents the clustering result

given a similarity matrix $M = \{ m_{ij} \mid 1 \leq i, j \leq n \}$ where $n$ is the number of students.
given an expected number of student groups, $G$, for heterogeneous grouping.
let triangular matrix $M' = \{ 100 - m_{ij} \mid m_{ij} \in M \text{ and } i < j \}$
let $\{s_1, s_2, s_3, \ldots, s_n\}$ denote the set of all students
$\forall i \in [1, n]$, let $C_i = \{ s_i \}$
let $C = \{ C_i \mid i \in [1, n] \}$
while $|C| > G$ and $\exists m_{ij} \in M'$ such that $m_{ij} \neq \text{NULL}$
   select an element $m_{ij} \in M'$ such that $m_{ij}$ is largest in $M'$
   find $C_h$ and $C_k$ in $C$ such that $s_i \in C_h$, $s_j \in C_k$
   if $|C_h| + |C_k| \leq (n / G)$
      $C_h \leftarrow C_h \cup C_k$
      remove element $C_k$ from $C$
      set $m_{ij} = \text{NULL}$
   end if
end while
if $\exists C_i \in C$ and $|C| \neq (n / G)$
   call `balance(C, n / G)`
end if
// $C$ represents the grouping result

Figure 6. Heterogeneous grouping algorithm with given number of groups
Step 1. Assign the expected number of groups, \( G \), according to the empirical rule or the learning strategy design.

Step 2. Due to heterogeneous grouping, first establish a triangular matrix \( M' \) as the difference matrix. The element value is 100 minus the element value of similarity matrix.

Step 3. Set every learner to be an independent group by default.

Step 4. If the quantity of all groups is currently larger than \( G \), select an element with the maximum value (the greatest difference) from all of the elements in the matrix \( M' \).

Step 5. Find the two learners (row and column of matrix \( M' \)) corresponding to the element value.

Step 6. If each of the found learners belongs to a different group, merge them into one group.

Step 7. Delete the maximum value element just selected. Repeat Steps 4-6 until the number of groups equals \( G \).

Step 8. Based on obtained heterogeneous grouping result, teachers can decide whether or not to set the difference threshold \( n_T \) of the number of people and execute the balance() function to conduct minor adjustments to balance the number of people.

Simulation and Experiment Evaluations

Simulation Assessment

This section conducts a simulation for evaluating and comparing the average intra-cluster diversity (AID) of clustering results generated by the proposed approach with random clustering, and clustering according to academic achievement. Formula (3) shows the calculation of the value of AID, which is the average of the differences of values of learners in each cluster. \( G_i \) represents the set of learners in the first cluster, while \( G \) represents the set of all clusters. \( m_{jk} \) is the value of similarity between the \( j \)th student and the \( k \)th student in the learner similarity matrix.

Because researchers try to achieve heterogeneous clustering, the goal of the simulation is to compare the difference between the results derived from three types of clustering methods and the learners in the average group. A higher value of AID implies greater heterogeneity, and greater heterogeneity implies better results of heterogeneous clustering.

\[
AID = \frac{\sum_{i=1}^{G} \left(\sum_{j=1}^{G} \sum_{k=1}^{G} |m_{jk}| \right) - m_{jk}}{\left(2 \left|G_i\right| - 2\right) \left|G\right|}
\]

Figure 7. Average inner-group diversity given a difference threshold of 70%

Figure 8. Average inner-group diversity given the respective number of 5
Figure 7 shows the average intra-cluster diversity (AID) of clusters generated with given a threshold of difference parameter of 70 percent. The label ‘Random’ indicates that random clustering was used to generate groups of learners. ‘Achievement’ indicates that learners were categorized into groups according to their academic marks. ‘U-portfolio’ indicates that the systematic method of clustering proposed in this paper, which utilizes the u-learning portfolios, was used. In this case, although the number of students increased, the AID values of the groups resulting from the proposed heterogeneous clustering were greater than the AID values of groups resulting from random clustering and clustering by academic achievement. The AID values of these clusters were also given a threshold of difference at 70 percent. The result of the simulation demonstrates that the proposed heterogeneous clustering with a given difference threshold obtained a higher average intra-cluster diversity compared to random clustering and clustering according to academic achievement. A higher AID value promotes the effect of intra-group learning.

Figure 8 shows the average intra-cluster diversity of generated clusters given the respective numbers of groups. In this simulation, each generated cluster contains five students and the AID values of the proposed method of clustering are greater than AID values resulting from other criteria for clustering (random and academic achievement). The results illustrate that the proposed method generates better heterogeneous clusters for facilitation of collaborative learning activities than the methods of random clustering and clustering according to academic achievement. Results also show that increase in the number of students implies increase in AID value.

**Experiment Assessment**

**Participants**

Participants included 114 fourth-graders from four classes in an elementary school in Taiwan. The students consisted of sixty-one males and fifty-three females between the ages of nine and ten years who voluntarily participated in this experiment. These four classes were randomly assigned four different grouping methods. One class was the control group $G_R$ with a random grouping method (27 students). Another class was the control group $G_A$ with an achievement grouping method (30 students), and each of the students was assigned to a heterogeneous group according to their school achievement. The third was the experimental group $G_T$ using the proposed grouping method with a given difference threshold (27 students). The fourth was the experiment group $G_N$ using the proposed grouping method with a given group number (30 students).

**Procedures**

The school year in Taiwan consists of two semesters. Students in the four classes used the U-Plant learning system for personalized outdoor learning during the first semester and the learning behavior of students was collected and recorded in the u-portfolio database. Because the students had previously used this learning system, researchers did not have to spend a significant amount of time familiarizing students with the operation of the learning system. The experiment was conducted in the second semester and was devoted to lessons on plant biology. The second semester of the school year consists of sixteen weeks minus midterm and final examination weeks. For the first four weeks, students participated in traditional classroom learning with teacher instruction in basic plant biology. In the fifth week, students were given quick training in the operation of the learning system. Students experiencing any operational problems during the training period were given technical assistance by the teacher. At the end of the fifth week, the four classes were randomly assigned the four methods of clustering. From the 6th to 14th week, all grouped students from each class attended three 40-min. lessons per week. During the first two 40-min lessons, each student received a handheld device equipped with Wi-Fi and RFID technology to carry out ubiquitous collaborative learning activities in the school botanical garden. In the last 40-min lesson, students provided feedback about their learning experience and the teacher discussed this feedback with students. Finally, in the 15th and 16th weeks, each group was asked to write a report on learning about plant biology. Each group shared their report with the other groups. To improve the student presentations, these presentations made up 25 percent of the final grade. All groups were asked to upload their final reports and the teacher bulletined the best one. The overall process of experimentation was videotaped to facilitate follow-up investigation and analysis of behavior.
Preliminary of Data Analysis

In our experiment design, four classes were respectively assigned four different grouping methods. The students in control group $G_R$ were randomly grouped into groups of three students. The size of the control group $G_A$, which was grouped based on school achievement, also assigned three students for each group to carry out outdoor collaborative learning. In experimental group $G_T$, the teacher assigned a 70 percent difference threshold to divide the class into nine groups with three members in each group. The teacher also assigned an expected grouping number six to the proposed grouping algorithm in which the experimental group $G_N$ generated five students for each group. During the experiment, the collected learning behavior with frequency and sequence data were analyzed using Lag Sequential Analyses (LSA) to find out differences in behavioral patterns according to the four learner grouping methods (Astous & Robillard, 2002; Bakeman & Gottman, 1997).

Experiment Results

This experiment evaluated whether the proposed grouping approach was effective and helpful for learners and whether the heterogeneous grouping method achieved directional and structural behavior patterns to affect traditional collaborative learning behaviors. The learning behavior collected during the experiment activities was classified into eight categories including moving (M), losing (L), referencing (R), observing (O), answering (A), interacting (I), completing (C), and taking note (N). The lag sequential analysis was used to analyze the frequencies and sequences of each student in each group, resulting in four groups shown in Table 3, 4, 5, and 6.

Table 3. Frequency transition table of the control group $G_R$

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>10</td>
<td>55</td>
<td>458</td>
<td>250</td>
<td>0</td>
<td>117</td>
<td>0</td>
<td>97</td>
<td>987</td>
</tr>
<tr>
<td>L</td>
<td>52</td>
<td>6</td>
<td>3</td>
<td>187</td>
<td>0</td>
<td>163</td>
<td>0</td>
<td>49</td>
<td>460</td>
</tr>
<tr>
<td>O</td>
<td>17</td>
<td>29</td>
<td>42</td>
<td>792</td>
<td>1058</td>
<td>367</td>
<td>332</td>
<td>321</td>
<td>2985</td>
</tr>
<tr>
<td>R</td>
<td>193</td>
<td>32</td>
<td>982</td>
<td>129</td>
<td>1177</td>
<td>433</td>
<td>433</td>
<td>455</td>
<td>4042</td>
</tr>
<tr>
<td>A</td>
<td>951</td>
<td>5</td>
<td>1169</td>
<td>453</td>
<td>47</td>
<td>388</td>
<td>957</td>
<td>234</td>
<td>4204</td>
</tr>
<tr>
<td>I</td>
<td>57</td>
<td>2</td>
<td>273</td>
<td>196</td>
<td>271</td>
<td>56</td>
<td>371</td>
<td>207</td>
<td>1433</td>
</tr>
<tr>
<td>C</td>
<td>974</td>
<td>3</td>
<td>61</td>
<td>5</td>
<td>129</td>
<td>0</td>
<td>183</td>
<td>79</td>
<td>1251</td>
</tr>
<tr>
<td>T</td>
<td>5</td>
<td>3</td>
<td>78</td>
<td>87</td>
<td>118</td>
<td>77</td>
<td>105</td>
<td>21</td>
<td>522</td>
</tr>
<tr>
<td>Total</td>
<td>2259</td>
<td>163</td>
<td>3066</td>
<td>2099</td>
<td>2698</td>
<td>1837</td>
<td>2198</td>
<td>1564</td>
<td>15884</td>
</tr>
</tbody>
</table>

Table 4. Frequency transition table of the control group $G_A$

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>12</td>
<td>46</td>
<td>573</td>
<td>352</td>
<td>0</td>
<td>254</td>
<td>0</td>
<td>105</td>
<td>1342</td>
</tr>
<tr>
<td>L</td>
<td>43</td>
<td>7</td>
<td>4</td>
<td>152</td>
<td>0</td>
<td>149</td>
<td>0</td>
<td>37</td>
<td>392</td>
</tr>
<tr>
<td>O</td>
<td>21</td>
<td>22</td>
<td>31</td>
<td>985</td>
<td>1123</td>
<td>784</td>
<td>689</td>
<td>357</td>
<td>4012</td>
</tr>
<tr>
<td>R</td>
<td>175</td>
<td>19</td>
<td>1065</td>
<td>132</td>
<td>1161</td>
<td>774</td>
<td>646</td>
<td>569</td>
<td>4541</td>
</tr>
<tr>
<td>A</td>
<td>1073</td>
<td>5</td>
<td>1226</td>
<td>386</td>
<td>31</td>
<td>412</td>
<td>1083</td>
<td>321</td>
<td>4537</td>
</tr>
<tr>
<td>I</td>
<td>63</td>
<td>2</td>
<td>684</td>
<td>217</td>
<td>862</td>
<td>42</td>
<td>389</td>
<td>254</td>
<td>2513</td>
</tr>
<tr>
<td>C</td>
<td>1064</td>
<td>5</td>
<td>52</td>
<td>11</td>
<td>0</td>
<td>153</td>
<td>0</td>
<td>89</td>
<td>1374</td>
</tr>
<tr>
<td>T</td>
<td>6</td>
<td>23</td>
<td>65</td>
<td>78</td>
<td>125</td>
<td>87</td>
<td>95</td>
<td>17</td>
<td>496</td>
</tr>
<tr>
<td>Total</td>
<td>2457</td>
<td>129</td>
<td>3700</td>
<td>2313</td>
<td>3302</td>
<td>2655</td>
<td>2902</td>
<td>1749</td>
<td>19207</td>
</tr>
</tbody>
</table>

Table 5. Frequency transition table of the experimental group $G_T$

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>17</td>
<td>48</td>
<td>628</td>
<td>387</td>
<td>0</td>
<td>683</td>
<td>0</td>
<td>179</td>
<td>1942</td>
</tr>
<tr>
<td>L</td>
<td>45</td>
<td>8</td>
<td>32</td>
<td>295</td>
<td>0</td>
<td>487</td>
<td>0</td>
<td>106</td>
<td>973</td>
</tr>
<tr>
<td>O</td>
<td>29</td>
<td>13</td>
<td>153</td>
<td>1185</td>
<td>1217</td>
<td>1012</td>
<td>1063</td>
<td>523</td>
<td>5195</td>
</tr>
<tr>
<td>R</td>
<td>473</td>
<td>27</td>
<td>1152</td>
<td>673</td>
<td>1279</td>
<td>1026</td>
<td>1173</td>
<td>686</td>
<td>6489</td>
</tr>
<tr>
<td>A</td>
<td>1154</td>
<td>6</td>
<td>1342</td>
<td>1053</td>
<td>218</td>
<td>1043</td>
<td>1365</td>
<td>451</td>
<td>6632</td>
</tr>
<tr>
<td>I</td>
<td>879</td>
<td>3</td>
<td>1056</td>
<td>986</td>
<td>1197</td>
<td>783</td>
<td>1168</td>
<td>657</td>
<td>6729</td>
</tr>
<tr>
<td>C</td>
<td>1379</td>
<td>7</td>
<td>172</td>
<td>159</td>
<td>0</td>
<td>849</td>
<td>0</td>
<td>174</td>
<td>2740</td>
</tr>
<tr>
<td>T</td>
<td>11</td>
<td>34</td>
<td>449</td>
<td>298</td>
<td>252</td>
<td>639</td>
<td>209</td>
<td>65</td>
<td>1557</td>
</tr>
<tr>
<td>Total</td>
<td>3987</td>
<td>146</td>
<td>4984</td>
<td>4936</td>
<td>4163</td>
<td>6522</td>
<td>4978</td>
<td>2841</td>
<td>32257</td>
</tr>
</tbody>
</table>
Lag sequential analysis indicates the degree of confidence with which one type of data influences the occurrence of another. In other words, LSA observes certain behavior patterns which occur immediately after another behavior occurs. Therefore, this study performed statistics for calculating the sequences of behaviors and frequencies of each sequence of pair behaviors which occurred during the whole learning process. In Tables 3, 4, 5, and 6, the value of each cell with row item \(x \in \{M, L, O, R, A, I, C, T\}\) and column item \(y \in \{M, L, O, R, A, I, C, T\}\) represents the frequency of occurrence of behavioral pair \(x\) and \(y\) where behavior \(y\) occurred immediately after behavior \(x\). In other words, \(x\) belongs to the set of behaviors recorded on the left of the table while \(y\) belongs to the set of behaviors recorded at top of the table. After calculating frequency transition tables from collected behaviors data, the subsequence process calculates the statistical significance of observed adjacent behaviors via utilizing the \(z\) score proposed by Allison and Liker (1982). The statistic formula of the \(z\) score follows:

\[
z = \frac{P_{B|A} - P_B}{\sqrt{\frac{P_A(1-P_A)(1-P_B)}{(n-k)P_A}}} \tag{4}\]

where \(P_{B|A}\) is the observed proportion of behavior \(B\) occurrences at lag \(k\) after behavior \(A\) occurs; \(P_A\) and \(P_B\) are the quantity of observed proportions of behavior \(A\) and \(B\) respectively; \(n\) is the sample size of behaviors in the sequence. Tables 7, 8, 9, and 10 show calculation results and present a matrix of \(z\) statistics computed for every pair of adjacent behaviors. A calculation result greater than \(+1.96\) indicates a statistically significant continuity level (\(p < .05\)) with a 95 percent level of confidence. In another words, a significant relationship exists between the two behaviors.

### Table 6. Frequency transition table of the experimental group \(G_X\)

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>7</td>
<td>34</td>
<td>472</td>
<td>317</td>
<td>0</td>
<td>774</td>
<td>0</td>
<td>116</td>
<td>1720</td>
</tr>
<tr>
<td>L</td>
<td>37</td>
<td>5</td>
<td>29</td>
<td>189</td>
<td>0</td>
<td>389</td>
<td>0</td>
<td>84</td>
<td>733</td>
</tr>
<tr>
<td>O</td>
<td>18</td>
<td>6</td>
<td>127</td>
<td>1056</td>
<td>1174</td>
<td>1276</td>
<td>942</td>
<td>5061</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>287</td>
<td>21</td>
<td>1103</td>
<td>389</td>
<td>1181</td>
<td>1292</td>
<td>1007</td>
<td>487</td>
<td>5767</td>
</tr>
<tr>
<td>A</td>
<td>1035</td>
<td>5</td>
<td>1247</td>
<td>897</td>
<td>196</td>
<td>1056</td>
<td>1174</td>
<td>122</td>
<td>6015</td>
</tr>
<tr>
<td>I</td>
<td>997</td>
<td>3</td>
<td>992</td>
<td>739</td>
<td>927</td>
<td>1011</td>
<td>859</td>
<td>318</td>
<td>5846</td>
</tr>
<tr>
<td>C</td>
<td>1174</td>
<td>5</td>
<td>78</td>
<td>45</td>
<td>0</td>
<td>1153</td>
<td>0</td>
<td>122</td>
<td>2577</td>
</tr>
<tr>
<td>T</td>
<td>8</td>
<td>24</td>
<td>427</td>
<td>258</td>
<td>231</td>
<td>893</td>
<td>199</td>
<td>57</td>
<td>2097</td>
</tr>
<tr>
<td>Total</td>
<td>3563</td>
<td>103</td>
<td>4475</td>
<td>3890</td>
<td>3709</td>
<td>8001</td>
<td>4163</td>
<td>1912</td>
<td>29816</td>
</tr>
</tbody>
</table>

### Table 7. Adjusted residuals table (z-scores) of the control group \(G_B\)

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-3.72</td>
<td>-1.72</td>
<td>2.23*</td>
<td>0.28</td>
<td>-3.47</td>
<td>0.29</td>
<td>-3.63</td>
<td>-0.16</td>
</tr>
<tr>
<td>L</td>
<td>-2.13</td>
<td>-2.42</td>
<td>-4.33</td>
<td>-0.11</td>
<td>-9.72</td>
<td>0.14</td>
<td>-10.40</td>
<td>-3.47</td>
</tr>
<tr>
<td>O</td>
<td>-1.97</td>
<td>-1.67</td>
<td>-1.42</td>
<td>8.26*</td>
<td>11.77*</td>
<td>1.42</td>
<td>0.97</td>
<td>0.78</td>
</tr>
<tr>
<td>R</td>
<td>0.23</td>
<td>-1.61</td>
<td>11.21*</td>
<td>-0.72</td>
<td>12.77*</td>
<td>4.97*</td>
<td>2.51*</td>
<td>5.13*</td>
</tr>
<tr>
<td>A</td>
<td>10.28*</td>
<td>-2.23</td>
<td>13.31*</td>
<td>2.85*</td>
<td>-1.90</td>
<td>1.89</td>
<td>9.87*</td>
<td>0.25</td>
</tr>
<tr>
<td>I</td>
<td>-1.39</td>
<td>-3.27</td>
<td>0.37</td>
<td>0.17</td>
<td>0.21</td>
<td>-0.73</td>
<td>1.78</td>
<td>0.19</td>
</tr>
<tr>
<td>C</td>
<td>10.97*</td>
<td>-2.97</td>
<td>-3.23</td>
<td>-2.64</td>
<td>-5.17</td>
<td>0.18</td>
<td>-8.26</td>
<td>-0.16</td>
</tr>
<tr>
<td>T</td>
<td>-2.13</td>
<td>-5.23</td>
<td>-0.92</td>
<td>-0.12</td>
<td>-0.23</td>
<td>-0.64</td>
<td>-0.19</td>
<td>-1.42</td>
</tr>
</tbody>
</table>

### Table 8. Adjusted residuals table (z-scores) of the control group \(G_A\)

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-2.16</td>
<td>-1.21</td>
<td>5.23*</td>
<td>1.73</td>
<td>-5.27</td>
<td>0.83</td>
<td>-7.67</td>
<td>0.17</td>
</tr>
<tr>
<td>L</td>
<td>-1.37</td>
<td>-2.67</td>
<td>-3.63</td>
<td>0.51</td>
<td>-10.62</td>
<td>0.46</td>
<td>-11.31</td>
<td>-1.64</td>
</tr>
<tr>
<td>O</td>
<td>-1.98</td>
<td>-1.85</td>
<td>-1.91</td>
<td>9.76*</td>
<td>12.47*</td>
<td>7.79*</td>
<td>6.85*</td>
<td>1.87</td>
</tr>
<tr>
<td>R</td>
<td>0.63</td>
<td>-0.72</td>
<td>11.26*</td>
<td>0.32</td>
<td>12.89*</td>
<td>7.17*</td>
<td>5.97*</td>
<td>4.69*</td>
</tr>
<tr>
<td>A</td>
<td>11.64*</td>
<td>-3.13</td>
<td>13.52*</td>
<td>2.16*</td>
<td>-1.76</td>
<td>3.73*</td>
<td>11.93*</td>
<td>1.31</td>
</tr>
<tr>
<td>I</td>
<td>-0.69</td>
<td>-4.71</td>
<td>6.42*</td>
<td>0.69</td>
<td>8.87*</td>
<td>-1.48</td>
<td>2.84*</td>
<td>0.97</td>
</tr>
<tr>
<td>C</td>
<td>10.97*</td>
<td>-3.47</td>
<td>-0.89</td>
<td>-2.34</td>
<td>-6.46</td>
<td>0.59</td>
<td>-9.23</td>
<td>-0.06</td>
</tr>
<tr>
<td>T</td>
<td>-2.84</td>
<td>-2.34</td>
<td>-0.58</td>
<td>-0.37</td>
<td>0.27</td>
<td>-0.18</td>
<td>0.12</td>
<td>-0.94</td>
</tr>
</tbody>
</table>
Based on the results of the above z statistics, this study converted the calculations into diagrams of behavioral relationships. Figures 9, 10, 11, and 12 are the behavioral diagrams of the four groups.

**Table 9. Adjusted residuals table (z-scores) of the experimental group $G_T$**

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-5.83</td>
<td>-0.82</td>
<td>4.21*</td>
<td>1.92</td>
<td>-9.97</td>
<td>5.97*</td>
<td>-10.42</td>
<td>0.72</td>
</tr>
<tr>
<td>L</td>
<td>-1.21</td>
<td>-8.19</td>
<td>-1.97</td>
<td>1.12</td>
<td>-13.83</td>
<td>3.23*</td>
<td>-14.72</td>
<td>0.13</td>
</tr>
<tr>
<td>O</td>
<td>-3.23</td>
<td>-6.85</td>
<td>0.34</td>
<td>12.26*</td>
<td>12.59*</td>
<td>9.21*</td>
<td>9.97*</td>
<td>3.72*</td>
</tr>
<tr>
<td>R</td>
<td>2.92*</td>
<td>-4.34</td>
<td>10.92*</td>
<td>3.34*</td>
<td>12.97*</td>
<td>9.34*</td>
<td>11.97*</td>
<td>6.21*</td>
</tr>
<tr>
<td>A</td>
<td>11.28*</td>
<td>-8.97</td>
<td>14.42*</td>
<td>9.64*</td>
<td>0.97</td>
<td>9.42*</td>
<td>14.52*</td>
<td>2.02*</td>
</tr>
<tr>
<td>C</td>
<td>13.29*</td>
<td>-8.28</td>
<td>0.48</td>
<td>-0.29</td>
<td>-11.28</td>
<td>7.26*</td>
<td>-12.92</td>
<td>0.59</td>
</tr>
<tr>
<td>T</td>
<td>-7.62</td>
<td>-2.42</td>
<td>2.31*</td>
<td>1.28</td>
<td>1.03</td>
<td>4.59*</td>
<td>0.92</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

**Table 10. Adjusted residuals table (z-scores) of the experimental group $G_N$**

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>L</th>
<th>O</th>
<th>R</th>
<th>A</th>
<th>I</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-6.52</td>
<td>-3.59</td>
<td>2.42*</td>
<td>1.21</td>
<td>-8.83</td>
<td>6.42*</td>
<td>-9.72</td>
<td>-1.05</td>
</tr>
<tr>
<td>L</td>
<td>-3.28</td>
<td>-7.15</td>
<td>-3.84</td>
<td>-0.41</td>
<td>-12.97</td>
<td>1.97*</td>
<td>-13.77</td>
<td>-1.59</td>
</tr>
<tr>
<td>O</td>
<td>-5.21</td>
<td>-6.88</td>
<td>-0.78</td>
<td>9.42*</td>
<td>11.28*</td>
<td>13.26*</td>
<td>7.97*</td>
<td>2.19*</td>
</tr>
<tr>
<td>R</td>
<td>0.85</td>
<td>-4.72</td>
<td>9.84*</td>
<td>1.90</td>
<td>11.85*</td>
<td>13.83*</td>
<td>8.63*</td>
<td>2.82*</td>
</tr>
<tr>
<td>A</td>
<td>9.17*</td>
<td>-7.93</td>
<td>12.97*</td>
<td>7.19*</td>
<td>-0.29</td>
<td>12.19*</td>
<td>10.34*</td>
<td>0.73</td>
</tr>
<tr>
<td>I</td>
<td>8.28*</td>
<td>-8.64</td>
<td>8.19*</td>
<td>5.90*</td>
<td>7.72*</td>
<td>8.79*</td>
<td>6.72*</td>
<td>1.59</td>
</tr>
<tr>
<td>C</td>
<td>10.84*</td>
<td>-7.49</td>
<td>1.89</td>
<td>-2.92</td>
<td>-10.15</td>
<td>10.15*</td>
<td>-11.40</td>
<td>-0.92</td>
</tr>
<tr>
<td>T</td>
<td>-6.26</td>
<td>-4.19</td>
<td>1.92</td>
<td>0.23</td>
<td>0.19</td>
<td>6.89*</td>
<td>-0.19</td>
<td>-2.74</td>
</tr>
</tbody>
</table>

**Figure 9. Behavioral transfer diagram of the control group $G_R$**

**Figure 10. Behavioral transfer diagram of the control group $G_A$**
Figure 9 shows a strong link in control group \( G_R \) between Answering (A) and Observing (O), indicating that most students observe empirical objects of learning to answer quizzes generated by the learning system. Referencing (R) and Answering (A) are also strongly linked, indicating that students frequently answered the quizzes by referring to the explanations, hints, or notes presented by the learning device. Referencing (R) and Observing (O) also have a strong relationship, indicating that students are more attentive to empirical objects of learning due to the use and assistance of the learning system. In addition, a behavior sequence of \( M \rightarrow (O \rightarrow R \rightarrow A) \) is shown in the diagram, and \( O \rightarrow R \rightarrow A \) formed a sequence cycle. These results showed that students moving to a correct learning location concentrated on observing the plant with materials or explanations provided by the system, and attempted to finish the tasks.

Compared to the control group \( G_R \), the behavioral pattern of control group \( G_A \) (achievement grouping) increased connections and interactions between learning behaviors (Figure 10). Behaviors of Observing (O) and Answering (A) in control group \( G_R \) have a stronger relationship with Interacting (I) than do Observing and Answering in control group \( G_A \), indicating that students discuss the learning materials when they observe the plants and answer the questions. The relationship between Answering (A) and Completing (C) in control group \( G_A \) is stronger than in control group \( G_R \), showing that the students have better comprehension and degree of completion of assignments when they are grouped using the basic heterogeneous clustering method with student achievement data. This behavioral pattern generates two cycles of \( O \rightarrow R \rightarrow A \) and \( O \rightarrow I \rightarrow A \), indicating that heterogeneous clustering and clustering according to academic achievement could not only enhance the learning interests of students, but also promote more discussion among students.
Figures 11 and 12 reveal that the behavioral patterns of the experimental groups $G_T$ (grouping with given difference threshold) and $G_N$ (grouping with given number of groups) were more complex than the two control groups, and behavior in the two experimental groups was more brisk and interactive than behavior in the two control groups. Because the learning behavior of learners in the previous u-learning empirical learning environment was considered for heterogeneous clustering, these learners had higher heterogeneity. Therefore, the difference between the learners was also more significant than the difference resulting from clustering in the control group. This significant difference indicated that members of the experimental group had more interactive discussion during the process of group learning and more exchange of opinion and cognitive communication. These results demonstrated that the method of heterogeneous clustering that considered students’ behavior in previous U-plant learning activities had more effect on the students’ level of cooperation and social interaction in work.

Figure 11 shows that almost all types of behavior were linked with other types of behavior. A possible inference is that many students with higher heterogeneity encourage each other to more actively acquire new skills, ideas and knowledge, and build solutions to educational problems by working together. A strong behavior sequence from $O \rightarrow A \rightarrow C \rightarrow M$ is shown, indicating that the majority of students exhibited greater degrees of accomplishment.

Although Figure 12 resembles Figure 11, comparison of the relationship between behaviors reveals a different intensity of connections between the learning behaviors, particularly connections to the Interacting (I) behavior. The links to Interacting (I) behavior in experimental group $G_N$ were stronger. This phenomenon indicates that much discussion and interaction occurred between the group members, possibly because the students were grouped according to their previous U-plant learning behavior, and the size of each learner group in $G_N$ was larger than the groups in $G_R$, $G_A$, and $G_T$. Due to the larger group size, the degree of mutual interaction required to reach a consensus on learning and seemed more intense and frequent during the learning activity. Although intense discussion could lead to deeper understanding among the students, it could also generate too many distractions that could hinder the accomplishment of tasks and purpose of learning. Therefore, in a u-learning environment with high heterogeneous clustering of learners, the quantity of learning groups and the number of people in each group must still be controlled to avoid an inability to achieve a consensus on learning and reach goals for group learning.

The results above show that the behavioral relationships generated from the two experimental groups are more complex than the behavioral relationships in the two control groups. These results demonstrate that groups with a greater diversity of behavior exhibited more interaction between learners and effected the process of learning more significantly.

Conclusion

A context-aware ubiquitous learning environment has useful functions for gathering data on the learning behavior of students. These u-portfolio data can help enhance the analysis of the behavior, habits, styles, capability, and potential of learners, which could improve learning designs. In addition, research suggests that collaborative learning with heterogeneous group composition positively affects learners in regards to positive interdependence, social skills, interaction, and better group performance.

This paper proposes a systematic process for analyzing u-portfolios, building portfolio grids, calculating a learner similarity matrix, and generating heterogeneous learner groups for collaborative learning. Moreover, this study attempts to investigate the effects of different methods of group composition that consider the early learning behavior of students, and explores whether the proposed method of clustering significantly influences the behavior of learners in ubiquitous collaborative learning. The first evaluation utilized the simulation technique to analyze the efficacy of intra-cluster diversity with different clustering methods. Results indicated that the AID values of the proposed clustering algorithms were significantly greater than the AID values of clusters produced by other methods of clustering. The second evaluation experimented to evaluate the proposed clustering methods and utilized Lag Sequential Analyses methodology to assess learning behaviors in the experiment. The results indicated that the proposed clustering algorithms generated highly interactive learning behavior. Future research could address further concerns such as utilizing the Fuzzy or Multi-Repertory Grid to improve results and thoroughly experimenting with and analyzing the effects on learning.
Acknowledgments

This work was supported in part by the National Science Council (NSC), Taiwan, ROC, under Grant NSC 97-2511-S-006-001-MY3, and NSC 98-2631-S-024-001.

References


Using Online EFL Interaction to Increase Confidence, Motivation, and Ability

Wen-chi Vivian Wu¹, Ling Ling Yen² and Michael Marek³

¹Providence University, Taichung, Taiwan 43301 // ²National Central University, Taoyuan County, Taiwan 32001 // ³Humanities 409A, Wayne State College, Wayne, Nebraska, 68787, USA // wcwu@pu.edu.tw // linglingyen38@gmail.com // mimarek1@wsc.edu

*Corresponding author

ABSTRACT

Teachers of English as a Foreign Language (EFL) in Taiwan often use an outdated lecture-memorization methodology resulting in low motivation, confidence, and ability on the part of students. Innovative educators are exploring use of technology, such as videoconferences with native speakers, to enrich the classroom; however few guidelines have been developed for effective videoconference instructional design. This study used a survey methodology, Exploratory Factor Analysis, and Structural Equation Modeling to examine which elements of learning via videoconferencing most beneficially affect motivation, confidence, and ability. The study found that long-term changes in ability are best predicted by enjoyment of the learning experience. The data also suggested that even a small amount of authentic interaction in English made students more comfortable in applying their skills, more confident in what they learned, and more inspired to make global, cross-cultural connections. Therefore, EFL instructors should strive to use student-centered active learning and to offer their students interactions with native speakers, including interactions via distance technology.

Keywords

Videoconferencing, Computer-mediated communication, EFL, Motivation, Confidence, Ability, Constructivism, Online

Introduction

Immediacy of communication is one of the hallmarks of the global society of the 21st century. Business, politics, and the media all demand and expect seamless international exchange of information and ideas, and English is often the language of international interaction (Su, 2006). When two people interact who are not native speakers of the same language, they are likely to find common ground in English. The result is that instruction of English as a Foreign Language (EFL) is now a global priority for economic development, science, and interaction among governments. But in spite of the emphasis in many countries on producing college graduates with English skills, instructional methodologies have not always kept pace with the requirements of the marketplace. In countries where there is not a surrounding population using English actively, the language is still often taught as a traditional classroom subject, with students rarely interacting with anyone except their teachers and classmates – far from an authentic learning environment. Today, however, technology provides a global infrastructure serving business, political, social, and entertainment endeavors. This provides many new potential channels for interaction among people who speak different languages, live in different countries, and reside in different cultures; however educators must be willing to take advantage of the potential to use such interaction as a learning tool.

In Taiwan, the setting of this study, most people have minimal need to speak English on a daily basis, so English is instructed as a foreign language (EFL) and learning happens without any immediate opportunity to use English for actual communicative functions (Lan, 2005). In addition, EFL teachers in Taiwan often continue to use outdated lecture/memorization methodologies. These environments rarely include meaningful interaction with native speakers of English or authentic materials that relate to the target culture (Su, 2008; You, 2003). The result is that students are often not internally, integratively, motivated to pursue their study of English, resulting in lower proficiency.

However, there are demonstrated benefits to be gained from authentic experiences related to the target language, especially conversation with native speakers (Fujii & Mackey, 2009; Gilmore, 2007). Creative teachers attempt to replicate the target language’s environment, usually through bilingual curricula, technology-assisted teaching, and immersion programs, thus injecting authenticity and shifting the focus of the classroom from lecture and memorization to active learning. Savignon (1998) pointed out years ago that the classroom context is always different from a natural learning environment, but also concluded that teaching for communicative competence should be the guiding principle of English pedagogy where learners expect and value communicative skills (Savignon & Wang, 2003). Reliance by instructors on lecture and rote memorization makes this goal difficult to achieve. Institutional culture, technology choices, characteristics of teachers and students, instructional design, and
pedagogic criteria can all affect the success of such efforts to foster active learning via authentic experiences, including those available via technology (Fresen, 2007).

Study framework

This paper reports on a project that used student-centered, active learning, and instructional materials that students viewed as highly authentic, including live online interaction with a native English speaker on topics of American culture. The American researcher spoke repeatedly via Internet videoconference to English conversation classes taught by the Taiwanese researcher. After short presentations by the American, the students each talked briefly with the American to ask questions or make comments about the subject of the presentation. This project was based on an extensive review of the relevant academic literature on the learning dimensions of motivation, confidence, and ability. These dimensions were selected because motivation and self-confidence are often described in the academic literature as predictors of academic performance, i.e. actual ability (Tavani & Losh, 2003).

Many studies have examined videoconferencing as an approach to EFL instruction. For example Wu and Marek, (2010) found that as a result of a series of videoconference instructional sessions, motivation, confidence, and ability of the students correlated directly and confidence profoundly impacted perceived ability in students. However, details are still lacking in the literature about the multiple components that make up these factors. Thus, the focus of current research was not simply to reexamine relationships among these three factors, but, further, to gain deeper understanding about how the components of the factors impact one another using exploratory, confirmatory factor analysis, and structural equation modeling (SEM). Therefore, the research method used in this study pioneers a new approach to data analysis of videoconferencing for EFL instruction.

Types of motivation

Motivation, confidence, and ability are interrelated and interact with each other (Butler & Lumpe, 2008; Phillips & Lindsay, 2006). Motivation can increase rapidly, given a positive stimulus, but ability improvement may take significant time and study. In addition, confidence is a reflection of the other two factors because confidence grows as student ability increases and anxiety decreases, thus stimulating both motivation and ability. All three learning variables — motivation, confidence, and ability — are the result of the cumulative experiences of the student, both in and out of the classroom. All three variables improve or decline as the consequence of positive or negative experiences that motivate or de-motivate the students (Sakai & Kikuchi, 2009). As a result, EFL teachers have the unique opportunity to improve student motivation through fostering desirable student goals, stimulating active learning, and leading dialog about the purposes of learning.

Intrinsically motivated learners have long been considered to be more successful because their learning goal is to achieve satisfaction and enjoyment (Wang, F. X., 2008). Learners driven by extrinsic motivation tend to make the minimum effort required to avoid punishment or to gain rewards. Gardner’s framework of Instrumental and Integrative motivation (2001) is commonly cited. EFL students who are instrumentally motivated are extrinsically driven, studying English only enough to complete a required class, to acquire minimum required job skills, or to earn a degree — all external, utilitarian goals (Wu, 2006). Gardner considered integrative motivation to be more desirable and effective because it stems from the learner’s intrinsic desire to engage with the target language and culture. Gardner concluded that integrative motivation is a strong predictor of success in learning a foreign language.

Changing realities of the 21st century have led to the understanding that EFL learners today do not use their language skills only to communicate with native speakers. They may, as often, communicate with other non-native speakers in English (Kormos & Csizér, 2008; Lamb, 2004). Dörnyei (2005) offered a complex understanding of motivation in which foreign language learners envision an idealized English-speaking self, based in part on real-life encounters with speakers of the target language and in part on how the students imagine themselves functioning in a cosmopolitan international society. The key to imagining this international society is knowing that students will interact in English with people who are native speakers of many other languages. Yashima, et al. also found that social interaction with other cultures promotes intrinsic motivation of students (Yashima, Zenuk-Nishide, Shimizu, 2004). However, even though learners in Taiwan are constantly exposed to a wide range of English products and artifacts, such as American films, music, books, and videos, direct contact with native speakers on the daily basis is
often minimal (Cheung, 2001) and many researchers have found that the primary cause of lower English proficiency among students in Taiwan is weak learning motivation stemming from passive learning environments (Wu & Ke, 2009).

**Motivation factors and perceived English ability and confidence**

Students’ confidence in language use is reflected in whether they are willing to communicate (Yashima, 2002; Yashima, Zenuk-Nishide, Shimizu, 2004). Students often decline to use English because they are embarrassed about their lack of fluency (Shamsudin, & Nesi, 2006), or because of conflicts and misunderstanding about the language and the culture (Muller-Hartmann, 2000).

More frequent intercultural contacts have been shown to increase self-confidence in the use of foreign languages (Clèment, Noels, & Deneault, 2001). When the contacts were positive and pleasant, the experience led students to interact more frequently in the foreign language both outside and inside the classroom. Because the experience was enjoyable, the increased self-confidence of the students, in turn, affected their motivation in a positive way. Therefore, successful interaction with native speakers can relieve student hesitancy to express themselves and increase their confidence in using the language.

**CMC learning**

Many researchers have found that use of computer-mediated communication (CMC) for language instruction benefited EFL students and that students perceived that their confidence in using English increased in a CMC environment (Ortega, 2009). Research has been conducted using asynchronous and synchronous CMC for L2 language learning (Liu & Chen, 2007, Payne & Ross, 2005; Tudini, 2003), and also for everyday decision-making (Smith, 2003, 2004). Asynchronous CMC benefits students by permitting delayed response to questions to allow careful construction of grammar (Hudson & Bruckman, 2002). The ideal synchronous CMC for L2 learning is speaking with a native speaker who provides good target language interaction.

Most college students experience online real-time interaction as part of their personal social networking and text messaging. The adaptation of social networking structures already in use to the EFL environment is a natural step (Campbell, 2004; Miyazoe & Anderson, 2009), but teachers usually lag well behind their students in use of technology, particularly Internet social networking systems.

**Videoconferencing for language instruction**

Developing online opportunities for language instruction poses important and difficult challenges (Wu & Bright, 2006). In American education, videoconferencing technology has been available for close to 35 years and online collaborative learning has become increasingly common and valuable (Kubasko, Jones, Trotter & Andre, 2008; Tiene & Ingram, 2001; Saw, Majid, Abdul Ghani, et al, 2008). Once available only at considerable expense, videoconferencing is now possible via the Internet with minimal expense and standard home consumer equipment (Dantas & Kemm, 2008). Without careful consideration of the instructional design, however, videoconference instruction tends to default to long-distance lecture with little interaction. As a result, students tend to be passive, as if they were watching television (Gillies, 2008). In Taiwan, few EFL faculty members are familiar enough with videoconference technology to employ it in the classroom.

**Study context**

The Taiwanese and American teachers in this study met as colleagues in an American Ed.D. program that stressed student-centered active learning. This literature review led them to the conclusion that online learning, used well, and properly managed by the teacher, providing authentic interaction with native English speakers, can be particularly well-suited to move learners from passivity into active, highly motivated learning. However the literature was lacking in guidance on how such a methodology affects variables in the learning process. The researchers, therefore,
concluded that a study providing guidance for instructional design and lesson plans would be highly relevant. In order to provide this relevancy, the main goals of this study were to determine (1) which elements of learning via videoconferencing cause the most beneficial changes in motivation, confidence, and ability, and (2) the degree to which motivation, confidence and ability change, as a result of the number of videoconference sessions.

Methods

Theoretical Framework

This study was informed by the social constructivist philosophy in which student collaboration and student/teacher interaction leads learners to evolve their own knowledge foundations, and which depends on the building of positive relationships and interaction among students and between students and instructors (Brandon & All, 2010). In this way, a learning community is formed that helps members construct new knowledge that is well-connected to other knowledge held by the students. Scaffoldin is a primary tool used by teachers to foster this interaction (Lee, 2003), a technique in which teachers initially give high support for the task to be accomplished, but slowly withdraw support to encourage the students to be independent.

Subjects

The subjects of this study were 227 non-major EFL learners from the business school of a technical university in central Taiwan, as part of the class requirements of five sections of a required English conversation class taught by the Taiwanese researcher. The learners included both day- and night-school students, traditional and non-traditional students, and roughly equal numbers of male and female students. The classes lasted for the entire academic year. Data were collected at the conclusion of both the fall and spring semesters. Over one academic year, students participated in five videoconferences with the American researcher, an American native English speaker. After an initial presentation by the native speaker on an American cultural topic, the students talked with the American or made presentations back on related topics. The Taiwanese teacher used scaffolding during the videoconferences, initially giving high levels of support during the interaction, and gradually withdrawing that support to encourage the students to become more and more independent in their interactions.

Survey instrument

The questionnaire used in this study consisted of four major sections based on Gardner and Lambert’s questionnaire (1972). Questions were added about student perceptions of the following topics:

- Intercultural learning (Yashima, Zenuk-Nishide, Shimizu, 2004),
- Learning English through different channels (Campell, 2004; Miyazoe & Anderson, 2009) related to the motivation variable,
- Confidence in making foreign contacts (Cheung, 2001),
- Learning western culture (Muller-Hartmann, 2000) related to the confidence variable,
- Speaking English accurately, and
- Understanding English conversation (Shamsudin & Nesi, 2006).

Because some new question items were added to Gardner’s survey, reliability and validity of the questionnaire were tested. An exploratory factor analysis was used to reduce some question items from the questionnaire and a confirmatory factor analysis was used to test construct validity of the questionnaire. The overall internal reliability was .92, with each section also scoring above .85, which is considered to have high reliability compared with the minimum Cronbach $\alpha$ of .75, which is considered reliable.

The 13 items in the section A asked about the degree of change in student interest and motivation in studying both the English language and the culture of the target language. The 10 items in the section B explored the students’ perceptions of change in their English-proficiency levels. The 11 questions in the section C asked about change in student confidence in using the language. The final section, D, asked for demographic information -- gender, age,
program type, years of English study, type of high school attended (technology- or academically-oriented), and experience with online learning and using technology. Each question used a five-point scale. The low end of the scale was labeled “significantly reduced” (= 1.00) and the high end of the scale was labeled “significantly increased” (= 5.00). The midpoint of the scale (=3.00) was labeled “no change.”

The Taiwanese researcher administered the survey, in Mandarin, to the students at the end of each semester in each of the five classes. The survey questions were carefully translated to preserve similar meaning in both Chinese and English.

**Exploratory Factor Analysis and Structural Equation Modeling**

In order to address the goals of this study, three paired t tests, ten exploratory factor analyses, a confirmatory factor analysis, and an SEM analysis were used to explore information about path correlations among the components of each variable. After the data were collected, and for purposes of the SEM analysis, the researchers hypothesized, based on the past research findings, that ability directly results from confidence, and motivation and confidence directly results from motivation (Hashimoto, 2002; Yashima, 2002; Tavani & Losh, 2003; Yashima, Zenuk-Nishide, & Shimizu, 2004). Figure 1 depicts the resulting hypothesized model that was tested.

![Figure 1. The hypothesized model for statistical testing](image)

**Findings**

The survey results were coded for analysis. In order to answer research question 1, about elements of learning via videoconferencing causing the most beneficial changes in motivation, confidence, and ability, the researchers used Exploratory Factor Analysis and Structural Equation Modeling (SEM). To answer research question 2, about the degree to which motivation, confidence and ability changed as a result of the number of videoconference sessions, the researchers used means and t-tests.

**Exploratory Factor Analysis**

The researchers used factor analysis to decrease the number of the components (factors) in each of three variables -- motivation, ability and confidence -- using a Maximum Likelihood method with varimax rotation. The analysis revealed an underlying pattern of relationships for each variable. A Kaiser-Meyer-Oklin (KMO) measure of the sample adequacy validated the fitness of the data for factor analysis, performed based on a factor loading of 0.5 or higher and an eigenvalue greater than 1. These findings are in line with Gorsuch (1983) who is commonly cited as noting that extracted variances of 40%–50% reflect an adequate factor structure for self-report scales.
Factor analysis of motivation and interest

The varimax rotation solution for Motivation (see Table 1) revealed that 37.54% of the variance was explained by the three factors, with component 1, **Motivation to learn English through different channels** contributing 15.31%; component 2, **Interest/enjoyment** contributing 12.02%; component 3, and **Intercultural learning** contributing 10.21%. Analysis of internal consistency reliability of these four components yielded a Cronbach $\alpha$ of 0.85, and the KMO of this analysis was 0.877.

<table>
<thead>
<tr>
<th>Component</th>
<th>% of Variance</th>
<th>Sample question item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Motivation to learn English through different channels</td>
<td>15.31%</td>
<td>Item 11: Your motivation to watch or listen to TV or radio programs in English?</td>
</tr>
<tr>
<td>2. Interest/Enjoyment</td>
<td>12.02%</td>
<td>Item 1: Your interest in your EFL class?</td>
</tr>
<tr>
<td>3. Intercultural learning</td>
<td>10.21%</td>
<td>Item 7: Your interest in making foreign friends?</td>
</tr>
<tr>
<td>Total</td>
<td>37.54%</td>
<td></td>
</tr>
</tbody>
</table>

Factor analysis of confidence

The two-component solution for Confidence (see Table 2) explained 47.38% of the variance, with component 1, **Confidence in learning English and western culture through videoconferences** contributing 31.04%; component 2, **Confidence in foreign contacts through videoconferences** contributing 16.34%. Analysis of internal consistency reliability of these three components yielded a Cronbach $\alpha$ of 0.89, and the KMO of this data analysis was 0.807.

<table>
<thead>
<tr>
<th>Component</th>
<th>% of Variance</th>
<th>Sample question item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Confidence in learning English and western culture through videoconferences</td>
<td>31.04%</td>
<td>Item C11: Confidence in studying Western culture?</td>
</tr>
<tr>
<td>2. Confidence in Foreign contacts through videoconferences</td>
<td>16.34%</td>
<td>Item C8: Confidence in learning English through distance learning?</td>
</tr>
<tr>
<td>Total</td>
<td>47.38%</td>
<td></td>
</tr>
</tbody>
</table>

Factor analysis of ability

The two-component solution for English Ability (see Table 3) explained 44.36% of the variance, with component 1, **Speaking accurate English** contributing 24.32%; component 2, **Understanding of English conversations** contributing 20.04%. Analysis of internal consistency reliability of these three components yielded a Cronbach $\alpha$ of 0.88, and the KMO of this data analysis was 0.900. Table 4 depicts significant intercorrelations among the components of the variables, motivation, confidence, and ability.

<table>
<thead>
<tr>
<th>Component</th>
<th>% of Variance</th>
<th>Sample question item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Speaking accurate English</td>
<td>24.32%</td>
<td>Item B8: Ability in pronouncing English words more accurately?</td>
</tr>
<tr>
<td>2. Understanding of English conversations</td>
<td>20.04%</td>
<td>Item B3: Ability in understanding English in conversations with Native speakers?</td>
</tr>
<tr>
<td>Total</td>
<td>44.36%</td>
<td></td>
</tr>
</tbody>
</table>
### Table 4. Intercorrelations among the components of the variable, motivation, confidence, and ability

<table>
<thead>
<tr>
<th></th>
<th>Motivation component 1</th>
<th>Motivation component 2</th>
<th>Motivation component 3</th>
<th>Confidence component 1</th>
<th>Confidence component 2</th>
<th>Ability component 1</th>
<th>Ability component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation component 1</td>
<td>1.00</td>
<td>0.487**</td>
<td>0.466**</td>
<td>0.434**</td>
<td>0.346**</td>
<td>0.288**</td>
<td>0.295**</td>
</tr>
<tr>
<td>Motivation component 2</td>
<td></td>
<td>1.00</td>
<td>0.507**</td>
<td>0.313**</td>
<td>0.189**</td>
<td>0.165*</td>
<td>0.250**</td>
</tr>
<tr>
<td>Motivation component 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.357**</td>
<td>0.276**</td>
<td>0.244**</td>
<td>0.377**</td>
</tr>
<tr>
<td>Confidence component 1</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.744**</td>
<td>0.616**</td>
<td>0.558**</td>
</tr>
<tr>
<td>Confidence component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.569**</td>
<td>0.494**</td>
</tr>
<tr>
<td>Ability component 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.678**</td>
</tr>
<tr>
<td>Ability component 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

See tables 1, 2 & 3 for identification of motivation, confidence and ability components.

![Figure 2. SEM result of the latent variables motivation, confidence, and ability](image)

### Structural Equation Modeling (SEM)

An SEM with Amos 17.0 was used to test the CFA models for the questionnaire’s construct validity. Two latent variables consisting of the ability to speak accurate English and the ability to understand English conversations were entered, with the result indicating that the model fitness to the data was good, Chi-Square = 38.77, GFI = .965, RMESA = .047. Motivation to learn English through different channels, motivation in intercultural learning, and interest/enjoyment were found to define the latent variables related to motivation, and the model fitness to the data
was good, Chi-Square = 43.69, GFI = .965, RMESA = .04. Confidence in learning English and western culture through videoconferences and confidence in foreign contacts through videoconferences were found to define the latent variables related to confidence, and the result indicated that the model fitness to the data was good, Chi-Square = 26.40, GFI = .976, RMESA = .008.

All aforementioned latent variables were included and tested in a sequence of models with standardized path coefficients using Amos 17.0. The result of the final model is shown in Figure 2. The model’s fitness to the data was good: Root Mean Square Error of Approximation (RMSEA) = .03, Goodness-of-Fit Index (GFI) = .94, Root Mean Square Residual (RMR) = .035, Chi-Square = 144.75, P >.05. The standardized regression weights in the Figure 2 represent the amount of change in the dependent variable that is attributable to a single standard deviation unit worth of change in the predictor variable.

The empirical findings indicated that confidence in interacting with foreigners through videoconferences is the direct result of motivation to engage in intercultural learning, ability to speak accurate English, and ability to understand English conversations. Confidence in learning English and western culture through videoconferences is the direct result of enjoyment, ability to speak accurate English, and ability to understand English conversations. Ability to understand English conversations is correlated to motivation to learn English through enjoyment, and motivation in intercultural learning. Ability to speak accurate English is correlated to enjoyment.

Changes in the three variables after more videoconferences

To determine the influence of the number of videoconferencing sessions on motivation, confidence, and ability, mean scores for each were computed at the end of each semester. Comparing student responses from the end of the first semester (after two videoconferencing sessions) with the end of the second semester (the same students in a total of five videoconferences over the two semesters), the participants perceived a moderate increase in motivation from the end of the first semester to the end of the second semester, with a mean score increasing from 3.80 to 4.09, and a t-test value of $t(226) = 4.16$, $p < .05$. More videoconferencing sessions, therefore, improved the student’s perceptions of their own motivation at a significant level. Additional videoconferencing sessions, however, did not change their perceived ability or confidence at a significant level after conducting t-tests.

Discussion

This study took place within a culture in which the daily use of English is rare, one of two typical English-language contexts found in the Asia-Pacific region. Like in Japan and Korea, use of English is uncommon in Taiwan. On the other hand, in countries such as Hong Kong, Singapore, and the Philippines, a significant base of native English speakers leads to more frequent use of English in everyday life. This social use, or non-use, of English results in significantly different mindsets about the learning of English, as shown in many studies over the years (Liu & Jiang, 2009; Sheen, 2004).

Relationships among the three learning variables

The findings with respect to research question one, about the most beneficial elements contributing to motivation, confidence and ability, show that there were multiple complex interlinked dimensions. The SEM analysis showed that confidence in using English stemmed from multiple subcomponents of ability (ability to speak accurate English, enjoyment, and ability to understand English conversations), as well as confidence in making foreign contacts. Ability to understanding English conversation correlated in both directions with Motivation to learn English through different channels, enjoyment, and motivation for intercultural learning, all subcomponents of motivation, and also influenced confidence in foreign contacts through videoconferencing. Changes in ability to speak English accurately, as perceived by the students, correlated in both directions with enjoyment, a subcomponent of motivation. Figure 3 shows a basic conceptual model of the various categories of correlation found in this study, with the arrow directions showing key SEM path directions and the width of the lines showing the relative statistically significant correlations.
As figure 3 makes clear, English ability is the most important factor impacted by the instructional design of this study. The majority of students whose English abilities were low felt that their English abilities rapidly increased when interacting with the American professor. It is likely that the interactive classroom activities helped the students learn vocabulary and structure of sentences used in conversational English. Because their English abilities were inadequate for correct conversation, the students’ beginning confidence levels were also low. The findings indicated that motivation is the indirect factor impacting on ability and confidence of the participating students. Therefore, the students initially had low motivation for interacting with the American professor because of the language barrier. As the videoconference lessons proceeded, student ability, confidence, and motivation all grew.

Previous research is clear that an instructional methodology stressing interaction as a tool for building confidence produces increases in student ability and that student centered, active learning, including videoconferences similar to those in this study, results in direct correlations of motivation, confidence, and ability of students, with confidence profoundly impacted the students’ perceptions of their own ability (Wu & Marek, 2010). The current study also found that enjoyment is a major factor impacting student learning over the long term. But this study also implied that the students perceived difficulty in making actual conversations with native speakers because their English abilities were inadequate. Therefore, regardless of past English training, EFL students with low skills may require specific conversational training. The result of a positive experience is that as students are drawn into using English as a real, communicative language, they became more proficient and developed a higher level of motivation. This positive experience, in turn, promoted the mental image that the student needs to be proficient in English in order to partake in international society (Dörnyei, 2005).

The current study has shown that well designed videoconferencing for interaction with native speakers, rich in authentic cultural information, does increase confidence and improve motivation, which in turn have the effect of strengthening ability, over the long term. When matched with more conventional classroom activities that directly relate to ability, the instructional design in this project hits all three of the “bases” by improving ability, confidence, and motivation. Therefore, as Blake held (2009), authentic learning is not simply the ability to talk with a native speaker, but rather an environment in which students have the support they need to make meaning for themselves, based on authentic source material.

The instructional design used in this study was grounded in constructivism (Brandon & All, 2010), a predominant model of educational psychology around the world. Constructivism holds that people construct meaning and learn most effectively by active participation in the learning process. According to Constructivism, passive classrooms, such as the lecture-memorization model, are not effective teaching tools. In order to apply Constructivist principles, the teachers’ pedagogical design made the classroom in this study highly interactive. Following the American’s presentations, providing new information, the students collaborated actively to develop presentations back to the American, selecting new information, synthesizing it, and fitting it in with what they already knew to make new mental connections. Therefore, this study stressed multiple dimensions of interaction for the students, including with the native speaker, with their own teacher, and with their peers, both in and out of class, in formal and informal settings, in accordance with Constructivist principles. Furthermore, the instructional design employed scaffolding, a technique studied widely and found to be valuable (Van de Pol, Vilman & Beishuizen, 2010). Active, successful interaction and effective communication in the target language, be it with other students, teachers, or speakers of the language from outside the classroom, has proven to be a valuable component of an instructional design that builds ability through strengthening motivation and confidence.

The findings of this study also showed that this application of CMC created a safe environment in which motivated students could engage in language learning. As Dörnyei predicted, the more the learners felt connected
internationally, the more likely they were to succeed in increasing their perceived confidence and ability. The learning environment used in this study facilitated intercultural learning by, in effect, establishing an international learning network.

Young people in Taiwan readily access American music, TV shows, computer games, and even American foods. However, the American culture itself is often poorly understood in Taiwan, because its elements are not translated into the language which people use on the daily basis, or because the culture is not truly represented in the music, video and other elements exported to Taiwan. In this study, American culture was directly introduced through the videoconferencing interactions. In students’ opinions, actual interactions with the American native speaker allowed students to experience American culture more intimately, providing a strong connection with their own lives.

Conclusions

This study used statistical tools rarely, if ever, used previously to evaluate student-centered active learning classrooms with authentic source materials. The data-based findings using these advanced analysis tools indicate that the most fundamental factor in elevating all three learning variables is enjoyment. This is also intuitive, because students who are bored or who do not see the value in a course will not apply themselves. Because student-centered active learning improves enjoyment, EFL instructors must make this a priority in their instructional design. As part of this, the findings indicate that teachers should strive to offer their students successful interactions with native speakers, or excellent speakers of English from any other culture, on topics of particular interest to the students. Such successful, and therefore enjoyable, interaction builds student motivation and eventually leads to improvements in ability and confidence. The real benefit of such an instructional design is not just making students more willing to participate in videoconferences, but rather making them more confident in every kind of interaction in English, and also improving their English ability. Any type of positive communicative experience in the target language or with the target culture will ultimately strengthen the confidence of students, enhance their motivation, and, in turn, improve their ability.

The significance of this study, and its contribution to the academic literature, is that it suggests that use of technology in EFL instruction should always stress interaction and active communication, using both formal and informal dialog, and with scenarios for interaction that are as authentic, frequent, and enjoyable as possible. This is in contrast to many past efforts to use videoconferencing for foreign language enrichment, which have too often been staged as one-time “stunts” that had little role to play in the overall lesson plan or the curriculum design of the semester (O’Dowd, 2005). The benefit of, and best practices for, use of CMC technology in EFL instruction is a vital issue today, and this study shows that well designed interaction via Internet videoconference offers learners a usable, familiar medium for real communication that builds the “habit” of using English regularly.

Limitations of this study include that it is based at one technical university in Taiwan and it may not be appropriate to generalize it to other populations of EFL students. It also examines one particular repeated use of videoconferencing as part of a larger student-centered instructional design. Teacher-centered approaches may produce different results. Furthermore, the data is based on student perceptions of personal motivation, confidence and ability. Future research could employ external validation of changes in these learning variables.

The authors recommend that teachers employ scaffolding as a vital part of online interaction, allowing teachers to lead the students to become more and more independent and thus more confident in their interactions. Furthermore, instructors should understand that motivations to study English for men and women and for students of different ages may be different and use appropriate approaches to trigger their respective motivations.

One goal of any academic program should be to provide a foundation from which students can further develop their own ability to adapt and continue learning on their own. This study shows that in EFL realms, technology makes it possible to provide opportunities more commonly found only when there is a surrounding population of native speakers, and thus helps transform traditionally passive learners into more engaged and interactive learners. The data shows that even a relatively small amount of positive authentic interaction in the target language made students more comfortable in applying their skills, more confident in what they learned, and more inspired to make global, cross-cultural connections. Therefore, this instructional design positively influenced what Dörnyei called the vision of self of the students, promoting the idea of being able to function in the cosmopolitan 21st century international culture and leading to stronger overall EFL motivation, confidence, and ability.
Acknowledgements

This project was partially supported by the Taiwanese National Science Council grant #NSC 98-2410-H-126-033.

References


Investigation of Organizational Interaction and Support in an NGO through Computer-Mediated Discussions

Yao-Jen Chang and Yao-Sheng Chang

Department of Electronic Engineering, Chung Yuan Christian University, Chung-Li, 320, Taiwan, R.O.C. // Institute of Business & Management, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. // yjchang@cyu.edu.tw // chang801@gmail.com

ABSTRACT
Discussion forums have been used to support organizational communication and they have become a candidate for study of organizational behaviors. However, online behaviors of NGOs have been insufficiently studied compared to those studies conducted in education and industries. Our empirical study examined how social workers in one NGO used an internal discussion forum to help them disseminate time-sensitive employment opportunities for people with mental impairments, provide space for collective expressions of support, and enable sharing of information and experiences. The main purpose of the paper is to study a multifaceted assessment methodology to assess online discussion forums in an NGO. In this paper, we carry out a systemic set of assessment methods including network analysis, category analysis, and content analysis for evaluating online social software, and particularly discussion forums. The methods are used empirically to assess patterns of information sharing, social network, and peer support in an NGO.

Keywords
Assessment, NGO, Social Software

Introduction
Information technology definitely changes the whole world, but not everyone has a fair access to it. “Poor service for the poor” is a well-known saying in the area of social welfare. We now can rephrase it to say “Poor technology for the poor”. The information technology used by Non-Government Organizations (NGOs) serving the underprivileged is far less advanced than that of profit-oriented organizations (Currion, 2006). It is true that many online services are for free nowadays and considered by many people easy to use. However, it is not always the case for NGOs. First of all, different demographic groups acquiring a new IT skill often go through a different learning curve than in business or academia. Although technology is there, it is in general necessary to help the NGOs and humanitarian agencies adapt technology to their needs (Chang et al., 2010). For this and other reasons, the online behaviors of NGOs have been less addressed in the literature compared to other sectors (Kay 2006; Hall & Davison 2007; da Cunha & Orlikowski 2008). We were interested in understanding how virtual communication spaces can be used within an NGO to address potential opportunities and threats to the success to its programs.

Discussion forums are widely available computer-supported communication technologies that facilitate virtual interaction on the Internet. The number of discussion forums continues to increase powered by the growth of the World Wide Web. For example, Google alone hosted 4.3 millions of discussion forums as of August 2009 in almost every language, region, and category. In this paper, we examined how the social workers used a discussion forum to help themselves deal with organizational communications that they perceived as essential to their job functions. We conducted a three year project and participated in supported employment programs for people with mental impairments. Supported Employment is a well-defined approach to helping people with mental illnesses find and keep competitive employment within their communities. Supported employment programs are staffed by employment specialists who have frequent meetings with treatment providers to integrate supported employment with mental health services. Crucial needs of online communication among the participating NGO were identified as follows:

1. Job opportunity sharing and paperwork reduction: The legacy database was designed for regulating job coaches (job titles for the social workers in our research). Therefore, data contributed from job coaches is isolated from access by each other. A data sharing mechanism is needed to enable job coaches to benefit from the information sharing. A platform for job coaches to share job opportunity and improve the matching process is desired.
2. Mutual support among job coaches: Most supported employment programs for the mentally ill persons are operated under rehabilitation hospitals. As non-medical professionals and contract workers in medical settings, job coaches often find themselves isolated in the organizations they serve. Low job recognition of the mentally impaired trainees on the competitive workplace is constant frustration job coaches have to live with, which
creates tension, insecurity, and emotional burden. Information technology can facilitate support mechanism by establishing a platform for information exchange, knowledge sharing and social support.

3. De-stigmatization of mental illness: The major barrier toward employment of mentally impaired persons is the stigma for mental illness which devalues them and thus deprives them from contributing to the society. Supported employment enables the public to witness the fact that mentally impaired persons can work and be useful to others. Such narratives need to be told, recorded, and circulated to debunk deconstruct the myth about mental impairment. In fact, the success or failure of the NGO depends on such public awareness of social inclusion programs. The collective action of telling these stories by job coaches is possible via the intervention of information technology.

In addition to the obvious uniqueness in de-stigmatization, timely job opportunity sharing to place people with mental impairments in competitive jobs is also critical to the organization. An examination of its effectiveness through analysis of on-line communication would be important for other NGOs providing supported employment services to consider new practices. Emotional support is also a unique subject because social workers pay frequent visits to potential employers, community training workshops, and family of people they take care of. There is little time left for them to meet their colleagues face-to-face and share the joys and tears. This poses the need of analysis of on-line communication to see to what extent emotional support has been achieved.

**Setting and the purpose of study**

To address crucial needs of online communication, several types of social software were compared before reaching a decision to adopt a software environment. The only option of computer-mediated communication tool was email before the setup of discussion forum. In fact, an internal only blog was launched first. However, it was soon abandoned in a month because of the difficulty of finding someone to assume the authoring responsibilities in a busy NGO. In February and March of 2007, we conducted two focus group meetings with service providers in the NGO. The purpose was to explore whether there were common grounds regarding the existing problems and core requirements with regard to IT systems. In the second focus group meeting, we briefed the practitioners for 1.5 hours about the best practices of nonprofit technology used by the world top ten NGOs. After the two meetings, the on-line forum was identified as a requirement that may change the organization. Because none of them had expressed any particular preference about the discussion forum, Google Groups, a modernized discussion forum which can display multimedia in line and allow web pages to be created right inside the discussion threads, was used on a regular basis for disseminating time-sensitive job opportunities to colleagues, providing space for collective expressions of affection, and enabling sharing of information and experience.

The aim of this research is to analyze the use of a discussion forum to facilitate communications among social workers within an NGO. The hypothesis tested is that a discussion forum for social interaction increases the productivity of these workers in terms of paper reduction, timely information dissemination, and emotional support. Participation in the social network increases information integration and interaction and thus portends an enhancement of productivity in terms of job placement for disabled individuals.

Assessing the information flows and functioning of the systems is important to organizational efficiency and social welfare. Social interaction is measured by user traffic, categories of worker interactions, structure of interactions, and peer support. Unlike previous works that used single measurement methods (Kay 2006; Hall & Davison 2007), this paper proposes an assessment framework based on triangulation of traffic metrics, network analysis, categorization, and content analysis. The contributions of the paper include 1) examining an institute (NGO) that has received little academic attention in the literature, 2) the use of multiple assessment techniques to test the hypothesis and 3) discussing experiences, benefits as well as difficulties of increasing organizational efficiency even with the implementation of social software. We hope the multifaceted assessment methodology applied in this paper can stimulate discussion among the researchers of social welfare as well as general organizational studies.

This paper is organized as follows. We first consider some of the research foundations on discussion forums, and particularly those methods examining the use of discussion forums in organizations. We then describe the setting of our research study and the methods we used to analyze the forum data. We next discuss our findings in terms of the proposed practices. We conclude the paper by examining the research implications of the use of discussion forums to facilitate NGO programs as a computer mediated communication tool.
Research foundations

Kay (2006) surveyed an extensive number of articles and proposed a comprehensive metric for evaluating discussion forums for educational purposes. The metric was based on system collected statistics of user traffic, such as visits, time stamps, posts, and responses. A contribution of our work is that we extend the scope of assessments by adding social network analysis, category analysis, and content analysis so that the integrated metric materializes to fit the settings and interests of organizations with a special focus on information sharing, quality of sharing, and peer support. The strategy for the use of multiple methods and data sources (Mathison 1988) is employed to produce an evidence base for this study. Category analysis is to quantify the interactions of online discussions by their category rather than by the whole. This gives us a deeper understanding of their online behaviors through finer granularity. Second, social network theory (Scott 2000; Newman 2003; Carrington 2005, Eds.) is applied in order to identify the underlying interaction structure. Social network theory is an instrument to measure the structure of an organization, a community or a society. Social network analysis (SNA) assumes that the attributes of individuals are less important than their relationships and ties with other actors within the network. SNA leaves less room for individual agency and focus on the structure of the network. Social networks have been used to examine how organizations interact with each other (Gómez et al. 2008). Content analysis (Kember et al. 1999; Swain 2006) is employed to evaluate three attributes of online peer support, namely, reflection, propositional stances, and affective tones (Hall & Davison 2007). The degree of reflection demonstrated one’s depth of thoughts in challenging each other’s statements and attempting to make critical analysis. The propositional stances indicate one’s position taken in an argument or a stand on a given issue among the responses and the author of the original entry. The affective tone in produced content measures the extent to which offered support or exhibited destructive linguistic behavior is demonstrated.

With participation of more than thirty students in three years, our research was conducted in the form of service learning, a pedagogy that integrates domain specific research and experiential learning for university students (Sigmon 1979; Sigmon 1994; Furco 1996). It is essential that curriculum of service learning encourages students to apply what they learn in the classroom to enable NGOs to achieve more. Thus student research and curriculum can be integrated and opportunities for students to participate in communities can be more meaningfully created. The theme of service learning often presents itself in the context of action research (Friere 1970; Agyris, Putnam, and Smith 1985; Cornwall and Jewkes 1995; Baskerville 1999), the role of which helps us try to examine a local problem with a local point of view. Action research emphasize creating communities of inquiry within communities of practice (Friedman, 2001). This means that both researchers and practitioners must redefine their roles and develop a set of common values, norms, terminology and procedures. One of the objectives of action research is to make the tacit knowledge of practitioners explicit so that it can be critically examined and possibly changed in order to understand the problems under scrutiny and implement interventions to solve the problem. Action research in essence integrates practical problem solving with theory building. Furthermore, the interventions in action research often involve researchers and participants in the position of equals. This gives us an in-depth perspective for observing online behaviors exhibited in online forums.

Assessing an online discussion forum

Research Settings

The setting for this research is a Taipei-based NGO with supported employment programs for people mental impairments. The NGO has no communications department or any intranet, like most of the NGOs in Taiwan. In April and May of 2007, we conducted two focus group meetings with service providers in the NGO. The purpose was to explore whether there were common grounds regarding the existing problems and core requirements with regard to IT systems. In the second focus group meeting, we briefed the practitioners for 1.5 hours about the best practices of nonprofit technology used in the world top ten NGOs. After the two meetings, the on-line forum was identified as a genuine requirement that may change the organization. Google Groups was used as an online, internal only forum. None of the NGO’s senior executives were invited to register for this forum. The participating job coaches and the only one supervisor felt that it was secure to post to the forum. Except for two job coaches who preferred to use real identities, all the other participants used aliases when they posted or replied to discussions. Email addresses and true identities were collected in an off-line address book during enrollment. Although most participants of the online forum use an alias, their true identities can be recovered by email addresses that appear in each posting.
The objective of the forum is to fulfill the three needs as stated in the introduction: (1) job opportunity sharing and paperwork reduction; (2) mutual support among job coaches; and (3) de-stigmatization of mental illness. The employees of the NGO were expected to participate by posting and reading voluntarily on a regular basis. Fifty-one users participated in the discussions in the observed periods from Apr. 2007 to Mar. 2009. The population size reflected that it was a small study within an organization of social workers who dedicated themselves to maximizing employment opportunities for individuals with mental impairments. The culture there was very team-centric because a successful placement of individuals with mental impairments involved social workers, employment specialists and treatment providers. Field use of the discussion forum over a period of two years generated the dataset that we collected as a basis for knowledge discovery.

Data collection

Access to the NGO’s discussion forum was granted to the authors at all times during the research. As a common practice of action research, the participating social workers recognized that the content of the discussion forum was to be used as data and studied by investigators. The primary data consisted of archived postings and replies, supplemented with key interviews. This study follows in the tradition of research studies that focus primarily on electronic data (Wasko & Faraj 2005). Six interviews were conducted after all the empirical data analysis was completed to help us interpret the outcome of the research. The interviews were conducted with two participants who were major posters to the discussion forum, and four participants who occasionally posted but “lurked” extensively on it. In addition, focus group and participatory observation (Spradley 1979, Emerson 1983, Kvale 1996; Baskerville1999) helped us to become oriented in the supported-employment NGO setting during the time period of research.

Overview of collected data

We evaluated the discussion forum according to the traffic-centric metric (Kay 2006), as shown in Table 1. It was observed that 44 members initiated at least one topic. The number of unique subjects was 409 which contained 950 postings, some of them being replies to particular subjects. One hundred and seventy of 409 topics received replies, which was 43.2%.

The NGO has a bulletin newsletter to distribute employment opportunities for job coaches to choose for their mentally ill trainees. Unfortunately, the newsletter is issued biweekly due to staff workload. Some available openings can become closed before the newsletter arrives at job coaches. The average 65.1 hours of message time span with SD=46.4 hours means time reduction in job bulletin delivery from an average of one week to an average of less than 3 days.

Although traffic analysis indicates activity intensity levels, it provides little information about the underlying social network, their relationships and ties with other members within the network. To gain further insights from the perspective of network structure, social network analysis (SNA) is applied in the next section.

### Table 1: Traffic Metrics of Discussion Board

<table>
<thead>
<tr>
<th>Index</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Members (as of Mar. 2009)</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Members who initiate topics</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Topics</td>
<td>409</td>
<td></td>
</tr>
<tr>
<td>Replies</td>
<td>541</td>
<td></td>
</tr>
<tr>
<td>Topics with replies</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>Total messages (topics+replies)</td>
<td>950</td>
<td></td>
</tr>
<tr>
<td>Time span per topic</td>
<td>65.1 hr</td>
<td>46.4 hr</td>
</tr>
</tbody>
</table>

Network Analysis

Network analysis (Scott 2000; Carrington 2005, Eds.; Gómez et al. 2008) focuses on how the structure of ties (or linkages) in a social network affects individuals and their relationships. Traditionally, socialization into norms is believed to determine human behavior in networked organizations, network analysis is to determine the extent to
which the structure and composition of ties affect norms. Therefore, network analysis gives a perspective of structural study into online interactions of the NGO we studied. Network analysis based on the replies between members is shown in Figure 1.

![Network of members whose sums of replies are larger than 4. Directed links represent replies between members and diameters of circles indicate the sum of replies a member writes to and receives from others.](image)

**Figure 1.** Network of members whose sums of replies are larger than 4. Directed links represent replies between members and diameters of circles indicate the sum of replies a member writes to and receives from others.

Replies are a form of interaction online communities can exhibit as a social network. Network analysis with respect to replies indicates the strength of interactions among members of a community and identifies clusters whose intra-interactions are stronger than interactions with members outside the cluster. Strength of interactions between members is measured according to the sum of replies a member writes to and receives from others. In Figure 1, the maximum sum is 204 while the smallest is 4. Five different sizes of circles are therefore used to stand for five classes of members whose sums of outgoing and incoming replies range from (1) 0–40 (2) 41–80 (3) 81–120 (4) 121–160, and (5) 161 and above. It is noted that member 1 and member 10 both have 9 members who write replies to them, indicating the highest centrality in the network.

![Network of members whose sums of replies are larger than 80.](image)

**Figure 2:** Network of members whose sums of replies are larger than 80

Cluster identification is made by removing members from the network if their sums of replies are no larger than 80, i.e. less than 4 per month in average. The result is shown in Figure 2 where we see a clique of size 5. There are
bilateral interactions between every two members except the directed links from member 10 to member 1, from member 10 to member 2, and from member 4 to member 10. Therefore, we see that the core members of the network themselves have quite strong connections with each other, which can be interpreted as a force to tightly knit all the members together to form the social network we studied here. The findings are consistent with a previous research result that has shown that participation in online social networks is often unevenly distributed (Butler, Sproull, Kiesler, & Kraut, 2002). There is often a core group of participants who contribute most actively to the messages and to the running of the online social network, a peripheral group of participants who occasionally contribute a response or comment but rarely participate in the running, and a marginal group (often referred to as “lurkers”) who read but almost never post messages, and do not participate at all in management.

Social networks have been used to examine how organizations interact with each other. Social network analysis (SNA) assumes that the attributes of individuals are less important than their relationships and ties with other actors within the network. In other words, SNA leaves less room for individual agency and focus on the structure of the network. Traditional social scientific studies assume that it is the attributes of individual actors that matter. In the next section, we examine three attributes of discussion threads based on content analysis.

Content Analysis

To further investigate job coaches’ online social behaviors, we examined the characteristics of the discussion content with regard to depth of sharing and peer support. In light of the work by Hall and Davison (2007), the research questions were (1) the degree of reflection job coaches demonstrated in challenging each other’s statements and attempting to make critical analysis, (2) the propositional stances job coaches took between general agreement or disagreement among the responses and the author of the original entry, and (3) the affective tone in produced content ranging from offered support to exhibited destructive linguistic behavior. To extract the data from the discussion forum, we used content analysis, a technique common in the assessment of online interactions and one which has been deployed in other studies of social media (Kember et al. 1999; Swain 2006; Hall & Davison 2007). To study the effects of social support among participants, the focus of content analysis was on the replies rather than the main entries themselves. The unit of analysis we used is a whole comment.

The content coding scheme used in this study is summarized in Table 2. According to this scheme, the content of replies was characterized by the three independent dimensions. For example, a disagreeing reply with reflective thoughts and positive tones is coded as RDP, which stands for “Reflective, Disagree, and Positive” in terms of reflection, stances, and tones. The intention of the analysis was to reveal any emergent correlations among three attributes of replies collected in the dataset.

The affective dimension was generally quite straightforward and lent itself to comparatively easy classification. The affective tone can be positive, neutral or negative depending on whether the content is presented with a supportive tone, in a hostile manner, or with a neutral affection. To study the propositional stance, probing the range of agreement and disagreement requires more careful judgment and involves the researchers reading between the lines. The boundary between reflection and non-reflection was somewhat blurred for some occurrences of replies. Therefore, basic rules of good practice in coding (Fielding, 2001) must be adhered to. When problems or ambiguity arose, the context of the original entry was checked and a comparison of other similar cases was made to resolve the coding issue. Two researchers conducted independent analysis on the same dataset and the results were cross-examined. If there was any inconsistency, the individual items were retrieved for discussion and recoded on the consensus of the two researchers. For example, the statement “I think your opinion is insightful. What the employers really want is productivity. Therefore, the pre-service training got to be up to a high standard.” is judged as reflective by one research but non-reflective by the other. If they disagree, we need a third independent judge. The degree of care enhanced the reliability of the coded output and confidence in the process of conducting statistical analysis towards research findings.

Table 2: The three dimensional coding scheme

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Code</th>
<th>Interpretation</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflection</td>
<td>C</td>
<td>Context-free</td>
<td>Replies made out of the context of the original entry</td>
</tr>
<tr>
<td>U</td>
<td>Non-reflective</td>
<td></td>
<td>Replies made without demonstrating perceivable reflection on the original entry</td>
</tr>
</tbody>
</table>
Reflective replies made with substantial reflection

Propositional stance
A Agree Replies in agreement of the stance of the original entry
M Mixed Replies showing agreements on some points of the original entry but disagreeing with others.
D Disagree Replies disagreeing with the original points

Affective tone
P Positive Replies with a supportive tone
E Even Replies with a neutral affection
N Negative Replies are hostile, destructive, etc.

Table 3: Reliability measurement (Kappa coefficients), N=541

<table>
<thead>
<tr>
<th></th>
<th>Kappa</th>
<th>Asymptotic Standard Error (ASE)</th>
<th>Test of H0: p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflection</td>
<td>0.8478</td>
<td>0.0331</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Propositional stance</td>
<td>0.7490</td>
<td>0.0366</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Affective tone</td>
<td>0.7369</td>
<td>0.0419</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

There were 541 comments rated by 2 researchers on the three dimensions: reflection, propositional stances, and affective tones. In order to assess the inter-rater consistency, we used SAS 9.0 to calculate Kappa coefficients. The statistical reliability evaluation results are shown in Table 3. The three dimensions display a Kappa coefficient of 0.8478, 0.7490, and 0.7369, respectively, indicating that the rating had attained a fairly high level of reliability and evidence validity was confirmed.

The attribute-by-attribute coding results were recorded in Table 4. We found that the discussion forum has demonstrated highly supportive online environments for the participating members. Replies with positive tones are as high as 94% and stances showing agreement are 92% across all the five categories in the small population we studied. The affective dimension is considered supportive and positive according to the data. Therefore, “emotional support” was achieved and evidence of the supportive culture was confirmed. In the propositional stance, thirty-seven of the 43 replies with disagreement or mixed responses actually came from the category of learning of IT skills. An investigation into the content reveal that some of the social workers do not trust technology as they witnessed misuse of technology on the less privileged people in their experiences of social services.

Table 4: Content analysis and statistics (N=541)

<table>
<thead>
<tr>
<th></th>
<th>Reflective</th>
<th>Propositional stance</th>
<th>Affective tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflective</td>
<td>125</td>
<td>23%</td>
<td>508</td>
</tr>
<tr>
<td>Non-reflective</td>
<td>383</td>
<td>71%</td>
<td>33</td>
</tr>
<tr>
<td>Context-free</td>
<td>33</td>
<td>6%</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>541</td>
<td>100%</td>
<td>541</td>
</tr>
</tbody>
</table>

Reflective replies are less than one-fourth. Although reflective expressions of thoughts can raise the levels of discourse quality, composing reflective messages usually takes more time and effort to think than those that are not reflective. The social workers unanimously experience tight schedules and this is not uncommon for employees in the ever-restructuring NGO sectors. Still, some are willing to contribute reflective thoughts for other practitioners to think deeper about field plans, actions, and experiences. Although the administration category receives the most numbers of original postings and replies, reflective replies are less than 1%. We see that 44% of reflective replies go to the “emotional support” category while another 38% go to the category of “Learning of IT skills.” We saw reflective empathy ranging from feeling a concern for other people that creates a desire to help them, experiencing emotions that match another person's emotions, removing painful emotions from another, knowing what the other person is thinking or feeling, and so on.

An interesting question is whether reflection relates to personal identities. If it does, then the sample is biased and the results observed in our study may not be reproducible elsewhere. Since the study does not use a random sample, an independence test is required in particular. Furthermore, this paper is based on one group with small population. Therefore, it is critical to examine whether the sample is biased. The statistical validation of data samples was not
used in previous studies of content analysis (Hall & Davison 2007). We study (1) the relation between persons and the number of reflective replies they compose, and (2) the relation between persons and the number of reflective replies they receive. For both investigations, there is no statistical difference according to the test of normality, i.e. Kolmogorov-Smirnov test, indicating a good fit ($p>0.1$). Therefore, there is no difference between the distribution of the data set and a normal one. Furthermore, peer support offered to colleagues in terms of affective tones and propositional stances does not show statistically significant interactions with personal identities. Again, it indicates that there is no difference between the distribution of the data set and a normal one when evaluating to what extent identities affect peer support. It is confirmed by the raw dataset with more than 90% of replies with positive tones and replies showing agreement. Therefore, the culture of the discussion forum is largely supportive with individuals maintaining positive affection and avoiding disagreement.

**Performance by Categories**

To gain deeper insights into social interaction, discernable difference among subject types of online discussions is investigated. According to subject categories, messages that we studied can be categorized into 5 types: (1) general information sharing, (2) job opportunities, (3) emotional support, (4) learning of IT skills, and (5) administration. Earlier we facilitated computer workshops for social workers and NGO staff to adapt technology such as GMail and Skype to their needs. Those workshops stimulated their interests in learning latest IT skills so we include “learning of IT skills” as one type of themes. Statistics of the five categories are summarized in Table 5, including number topic items, number of replies, number of topics with replies, elapsed time between original postings and the first replies, if any, (response time), and elapsed time between original postings and the last replies, if any, (lifespan).

The administration category receives the most numbers of original postings and replies. It reveals that the NGO uses the discussion forum in their daily administrative matters most. It can be effective in paperwork reduction and information dissemination efficiency. The number of topics about learning of IT skills is the second highest. It is somewhat to our surprise that social workers, who didn’t receive much IT training in their careers, exhibit substantial interests in learning the latest trends of information technology and applications. Topics in the category of emotional support receive the highest percentage of replies. The discussion forum has been perceived as a reciprocal channel of showing concerns, support and care among the members.

<table>
<thead>
<tr>
<th>Category</th>
<th>Postings</th>
<th>Replies</th>
<th>Postings</th>
<th>%</th>
<th>Response with replies</th>
<th>Life span (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General information sharing</td>
<td>55</td>
<td>29</td>
<td>15</td>
<td>36</td>
<td>2866</td>
<td>4231</td>
</tr>
<tr>
<td>Job openings</td>
<td>77</td>
<td>99</td>
<td>33</td>
<td>52</td>
<td>22</td>
<td>3422</td>
</tr>
<tr>
<td>Emotional support</td>
<td>74</td>
<td>88</td>
<td>44</td>
<td>60</td>
<td>2712</td>
<td>5620</td>
</tr>
<tr>
<td>Learning of IT skills</td>
<td>81</td>
<td>133</td>
<td>37</td>
<td>42</td>
<td>3994</td>
<td>7430</td>
</tr>
<tr>
<td>Administration</td>
<td>122</td>
<td>232</td>
<td>63</td>
<td>51</td>
<td>637</td>
<td>7907</td>
</tr>
</tbody>
</table>

**Discussions**

The results of the study related to the functioning of the NGO in terms of paper reduction, peer support, and timely information dissemination, and particularly increased rates of job interviews. Table 6 summarizes main findings discovered in each method used in our study. The setup of such social software systems was quite straightforward and of minimal cost. However, to be successful an NGO needs to implement policies about online social interaction. Specifically, privacy issues need to be concerned with. The organization should make a clear statement on what could be written down on the system and what should not. When should the alternative communication tool be used that can help the system functioning better with privacy concerns? Information overload is an increasing problem in most workplaces. In fact, the interviews revealed that receiving irrelevant emails frequently was a major driver for the social workers to adopt discussion forums. NGO should encourage their staff to discern among the multiple communication tools existing simultaneously within the organization. For example, people use email frequently
when in fact blogs or forums should be considered in some occasions, such as making announcements or disseminating information to a large group of recipients. Careful use of communication tool will generally cut back the overload. These are among the issues to be considered in policy making when introducing new communication tools to organizations. Research findings also show that discussion forums are one of the appropriate tools in terms of peer support. However, communications in virtual spaces are only a small part of their life styles. There are still many occasions of emotional support that have to be delivered face-to-face or at least over the phone.

Table 6: Main findings in each method used in our study

<table>
<thead>
<tr>
<th>Data</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Analysis</td>
<td>activity intensity levels of online discussion</td>
</tr>
<tr>
<td>Social Network</td>
<td>a core group, a peripheral group, and a marginal group</td>
</tr>
<tr>
<td>Content Analysis</td>
<td>the discussion forum is highly supportive; some replies are reflective</td>
</tr>
<tr>
<td>Category Analysis</td>
<td>the discussion forum has been perceived as a reciprocal channel of showing concerns, support and care</td>
</tr>
</tbody>
</table>

In the beginning, we found the category of job opportunities exhibited the lowest rate of replies, a fact that was a little disappointing. To find out the reason, we conducted a series of interviews with the social workers, namely, the job coaches. They were excited when they were enabled to receive new job opportunities in the discussion forum. The visit frequency of messages in this category confirmed the enthusiasm at the early stages when the discussion forum was newly launched. It turned out gradually that the job openings flowing from the Council of Labor Affairs into the system were not categorized properly and therefore rarely prioritized persons with mental disabilities. Furthermore, most of the leads that appeared useful did not result in successful job interviews for their trainees. Therefore, the job category became less effective over time. The other reason why there was the lowest number of reply rates was due to ethics issues. Discussion of details about job openings may often involve sensitive issues and therefore the social workers would rather conduct the discussion over the phone in private.

To overcome the usefulness issue, the NGO changed the source of job openings to a clearinghouse maintained by an alliance of NGOs where job openings were better organized. Since job coaches were hesitant to discuss job openings for their disabled clients over the system because of privacy concerns, a Skype component was embedded by the system to each posting of job openings so that they could choose to talk over the phone instead of writing down text. The Skype use was considered a reply and counted by the system. Later, the rate of replies on the job opening category rose to the second place. Due to the timely dissemination of useful job openings facilitated by the system, the number of interviews almost doubled a year after the introduction of the system. The job coaches did not think the discussion forum took extra efforts due to two reasons. First of all, the information service contributed to achieving the objectives of the NGO in a positive direction. Second, the use of RSS syndication in discussion forums appeared easy to use and working for them. They didn’t bother to check in the discussion forum several times a day on the desktop or mobile phones; they simply let the RSS notify them when job openings showed up.

Limitations and future work

The study was based on a sample of 51 job coaches. First of all, the number of job coaches dedicated to supported employment for persons with mental disabilities in Taipei City (population 2.5 millions) has been below 100 in recent years. Because of the small population, a Kolmogorov-Smirnov test was applied in order to justify the outcome of statistical analysis. Second, access to content of NGO internal communications, especially those involved in providing services for persons with mental impairments, is very limited. It would be interesting to apply the set of assessment methods in investigating other important NGO areas with larger sample sizes. However, content analysis is labor intensive and its scalability to large amount of content is challenging.

Conclusions

Traditional site traffic analysis based on site visits, page views, number of posts, and length of stay provides only limited insights into human communication in the NGO of our case. We developed a comprehensive assessment method of gauging social software performance by network, category and content analysis. Network analysis
measures the connectedness of the social network. Performance by category reveals the most effective and ineffective subjects that users participate in. Content analysis helps implicit organizational culture surface and emerge. The integrated assessment gives a more comprehensive perspective of the organizational behavior in the social service group we studied.

Research implications of the use of discussion forums to facilitate NGO programs as a computer mediated communication tool are mixed. Given limited resources in most NGOs, computer mediated communication can be beneficial to paper reduction, timely information dissemination, and emotional support. In addition, insufficient human resource often puts NGO staff in extreme work burden, which may force them to diminish the use of computer mediated communication as it is often interpreted as extra efforts. Furthermore, the needs of NGOs have been largely underserved by the ICT industries compared to other sectors. To change their attitudes towards computer use, the information service should prove relevant to their job functions and meet the objectives of the NGO in a positive direction. Inviting them to participating in the decision process, instead of just in the test stage at most, is also critical. The respects they received and felt through participation help build mutual trust between researchers and practitioners as well as change attitudes towards a new information service. This can further decrease the level of resistance to familiarizing themselves with computer skills which, paradoxically, may be of help to the individuals and organization as well in the end.

References


A Simulation Model that Decreases Faculty Concerns about Adopting Web-Based Instruction

Hae-Deok Song¹, Wei-Tsong Wang² and Chao-Yueh Liu³

¹Department of Education, Chung-Ang University, Seoul, Korea // ²Department of Industrial and Information Management, National Cheng Kung University, Tainan City, Taiwan // ³Department of Public Administration & Policy, State University of New York at Albany, Albany, USA // hsong@cau.ac.kr

ABSTRACT

Faculty members have different concerns as they integrate new technology into their teaching practices. The integration of Web-Based Instruction in higher-education settings will not be successful if these faculty concerns are not addressed. Four main stages of faculty concern (information, personal, management, and impact) were identified based on Hall’s concern-based adoption model. By reviewing the literature on the diffusion of on-line education, the author identified support factors that may decrease faculty concerns about adopting WBI. System dynamics was used to examine associations between faculty concerns and these support factors. Based on these associations, a simulation model was built using STELLA to test the potential impact of support factors on the adoption of WBI by faculty members. The simulation model will aid educators and administrators in evaluating the impacts of support factors on the adoption of WBI.

Keywords

STELLA software, Adoption of WBI, Supporting factors for adoption of technology, Faculty concern, System dynamics

Introduction

Web-Based Instruction (WBI) has experienced explosive growth, and its use is becoming more attractive in higher education settings. Many innovation models have been proposed to provide a theoretical framework to facilitate the adoption of a new technology (Fullan, 2007; Rogers, 2003). An important point of view in the models is a person-centered approach (Derntl & Motschnig-Pitrik, 2005). According to Emrick, Peterson, and Agawala-Rogers (1977), two parallel dimensions exist simultaneously in the change process: 1) a systemic dimension that involves change in the environment of the user, and 2) a personal dimension, including cognitive, behavioral, and affective components, that involves the process of change within the individual. The system-centered approach clarifies essential factors that lead to technology integration changes processes (Hsu & Sharma, 2008). However, a common limitation of this approach is that it fails to look at the psychology of the innovation and, thus, the interventions are not persuasive enough to bring about the desired change. Research is needed to identify the personal-dimension variables that affect the adoption of WBI by faculty members (Ertmer, 2005; Georgina & Olson, 2008).

The concerns that faculty members have when deciding whether to integrate new technology are a critical condition that needs to be considered along with other personal dimension variables for the successful adoption of WBI in higher education settings (Adams, 2003; Matthew, Parker, & Wilkinson, 1998; Sahin & Thompson, 2007). The more concerns they have, the more likely they will be resistant to adopt the WBI. For this reason, it is important to identify the factors that can diminish faculty concern about adopting WBI. It is difficult and costly to test the relevant variables in practice, but simulating the impact of the model should allow educators and decision makers to assess the effectiveness of factors that may support the implementation of WBI in educational settings.

The purpose of this article is to propose a simulation model designed to test the impacts of the factors that support faculty adoption of WBI integration. In order to achieve this purpose, the stages of concern of the faculty members were identified using the Concern-Based Adoption Model (CBAM) (Hall, George, & Rutherford, 1977). The factors that support faculty WBI adoption regardless of the stage of concern were then suggested. These identified factors were based on a review of the literature on the diffusion of online education. System dynamics was used to determine associations between faculty concerns and these support factors. Finally, based on the identified factors that occur during the different stages of concern, a simulation model with examples of its use is presented.
Factors that decrease faculty concerns about adopting WBI

Faculty concern about integrating WBI

What concerns do faculty members have as they integrate new technology into their courses? According to Fuller (1969), the process of diffusion can be explained in terms of a psychological shift from the properties of an innovation to the concerns of the users. He initially proposed a model that described three phases of concern: a pre-teaching phase, an early teaching phase, and a late teaching phase. These three phases were later named "self," "task," and "impact" concerns, respectively. Hall, George, & Rutherford (1977) expanded upon these three stages of concern to end up with a total of seven stages. According to these stages, adopters advance from lower-level, self-oriented concerns (awareness, informational, and personal) to intermediate-level, task-related concerns (management), and finally to impact concerns (consequence, collaboration, and refocusing). In the awareness stage (Stage 0), a person has either little knowledge of or little involvement with the innovation. Self concern refers to the questions we ask when we hear about an innovation (Stage 1, informational) and think about how the innovation may affect us (Stage 2, personal). Task concerns emerge as we learn new skills such as time management and material usage (Stage 3, management). Impact concerns describe our thoughts on how we can make an innovation work better for our students (Stage 4, consequence), how to make it work better by actively improving it with colleagues (Stage 5, collaboration), and, ultimately, how to be successful with the innovation and seek out positive changes to implement (Stage 6, refocusing).

WBI, for the purpose of this paper, is the use of Web-based computer devices such as desktop computers, laptops, handheld computers, software, the Internet, and Learning Management Systems for instructional purposes (Hew & Brush, 2007). It is defined as a type of blended learning in which a significant portion of instructional activities is delivered online but traditional classroom instruction is not eliminated (Garnham & Kaleta, 2002). Most faculty members are expected to have concerns when integrating WBI into their teaching. However, these concerns will be different according to the stage of the adoption process they find themselves in. Rogers (2003) presented the steps in the innovation process as follows: knowledge, persuasion, decision, implementation, and confirmation. In general, when people are confronted with a new technology, they will gather information, test the technology, and then consider whether the new technology is a sufficient improvement to warrant the investment of their time and energy to learn the skills required to use it (Rogers, 2003). Based on Roger’s innovation process, five different categories of faculty members are assumed to appear in the WBI adoption process: 1) Faculty Unaware, who have little interests in the adoption of WBI, 2) Faculty Aware, who are aware of WBI and gather information about it, 3) Faculty Adopters, who apply WBI to their teaching, 4) Faculty Implementers, who regularly use WBI, and 5) Faculty Integrators, who are interested in extending the use of WBI in their teaching practices.

Based on Hall’s (1978) concern-based adoption model, faculty members are expected to experience different concerns during the process of innovation: awareness concerns for Faculty Unaware, information concerns for Faculty Aware, personal concerns for Faculty Adopters, management concerns for Faculty Implementers, and impact concerns for Faculty Integrators. Hall’s last three stages of concern (consequence, collaboration, and refocusing) were combined into one concern, called the impact concern, in this study because faculty members experience all of these concerns after adopting the technology into their classrooms.

Support factors that decrease faculty concerns about adopting WBI

How can we reduce these WBI-related concerns? Previous studies have suggested various factors that may contribute to technology adoption (Inan & Lowther, 2010; Ngai, Poon, & Chan, 2007; Sahin & Thompson; 2007, Selim, 2007, Wang & Wang, 2009). These factors include staff development opportunities, time, prompt technical, incentives and positive attitudes towards the technology (Buckenmeyer, 2001), improved student learning, equipment availability (Hew & Brush, 2007), ease of use, time needed to learn the skills required to implement the new technology, compatibility with materials, training, administrative support, personal comfort and colleague use (Olapiriyakul & Scher, 2006), perceived value, available resources and communication with other adopters (Keengwe, 2007), mission statements and institutional culture, faculty development programs (Kahn & Pred, 2000), personal conviction, motivation and experience, and organizational support. According to Bradshaw (2002), different support factors are needed based on the faculty member’s stage of concern. For instance, faculty members that are concerned with information need to know more about the technology. In this study, therefore, the support factors were
categorized according to faculty stage of concern. The following table summarizes the support factors and details the strategies that may decrease each type of concern (Table 1).

<table>
<thead>
<tr>
<th>Faculty Concerns</th>
<th>Support Factors</th>
<th>Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Faculty Unaware</strong> (Information Concerns)</td>
<td>Basic training</td>
<td>Basic application software, use of the Internet resources, email</td>
</tr>
<tr>
<td></td>
<td>Technology support</td>
<td>Support for hardware and software, access to technical staff, and other</td>
</tr>
<tr>
<td></td>
<td></td>
<td>equipment availability</td>
</tr>
<tr>
<td><strong>Faculty Aware</strong> (Personal Concerns)</td>
<td>Intermediate training</td>
<td>Effective use of web-based technology in the classroom, use of course</td>
</tr>
<tr>
<td></td>
<td>Instructional support</td>
<td>management software, and troubleshooting</td>
</tr>
<tr>
<td><strong>Faculty Adopter</strong> (Management Concerns)</td>
<td>Peer tutoring</td>
<td>Working with faculty mentors on their projects on an as-needed basis</td>
</tr>
<tr>
<td></td>
<td>Faculty incentives</td>
<td>Funding for technology purchases (hardware and software), financial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>compensation</td>
</tr>
<tr>
<td><strong>Faculty Implementer</strong> (Impact Concerns)</td>
<td>Advanced training</td>
<td>Assessment training (how to analyze student performance and evaluation</td>
</tr>
<tr>
<td></td>
<td>Administrative support</td>
<td>strategies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Institutional climate for technology, providing external motivation (e.g.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>promotion or releasing a course)</td>
</tr>
</tbody>
</table>

Faculty members with information concerns should be given entry-level information via available media, individually or in groups. At this stage, faculty members need basic training and technology support. Basic training includes downloading presentation software and learning how to develop simple presentations, and use Internet resources, e-mail, and simple course management software features, such as creating a syllabus. Technology support is also important to reduce this type of concern, which includes hardware and software support, access to technical staff for Web-based course development and Web-page development, and personal computers equipped with the required software applications (Hew & Brush, 2007; Zhao, Pugh, Sheldon, & Byers, 2002).

Faculty members with personal concerns should be granted the opportunity and encouragement to learn and talk about the technology and how to best use the technology. At this stage, the faculty members are in learning mode and need intermediate training and instructional support. Intermediate training must have an instructional focus that first guides faculty members to think about their teaching styles and then helps them decide how to integrate the web-based technology into their style of teaching. They must understand how technology can support their educational objectives. The training should help faculty members effectively use web-based technology in the classroom, incorporate creativity in their presentations, use course-management software, and troubleshoot. In addition to training, instructional support is needed; faculty members work with an instructional designer to learn more about the instructional use of technology and to design, develop, and evaluate technological applications.

Faculty members with management concerns need practical help in developing and implementing web-based activities. The strategies used to reduce the management concerns include peer tutoring, administrative support and incentives. Peer tutoring involves faculty mentors working individually with those seeking help on their projects on an as-needed basis. This encourages faculty members to “share expertise, perspectives and strategies with each other” (Newcombe & Kinslow, 2000). Faculty incentives include funding for technology purchases, providing promotion/tenure as external motivation for faculty members who integrate new technology methods, and time to a) develop and maintain the web-based technology, b) learn more about the technology, and c) attend training sessions and any other support activities. Lack of administrative support is often cited by faculty members as a key barrier to adopting new technology (Dooley & Murphrey, 2000). Administrative support involves creating an institutional climate conducive to technology use, and providing equipment.

Finally, faculty members who have impact concerns need to be involved in envisioning and planning the use of the technology (Bradshaw, 2002). The strategies used to reduce impact concerns should focus on decreasing all three types: consequence, collaboration, and refocusing. They consist of advanced training and administrative support. Advanced training aims to provide information on analyzing students’ performance and on evaluating different strategies. Administrative support should provide opportunities for faculty members to work together each other so
that they can share the knowledge and skills that they have learned, and should improve the quality of instructional and administrative support (Keengwe, 2007).

**Modeling the WBI adoption process with system dynamics methodology**

In this study, a system dynamics approach was adopted to investigate the behaviors of faculty members as they adopt WBI. System dynamics is a method used to enhance the learning of complex social systems by helping interested persons to learn about the complexities involved, understand the sources of resistance to policies, and design more effective policies (Sterman, 2000).

An important step towards understanding a complex social system is determining policy structure. A policy structure needs to be created to represent the complexity of a certain system (Richardson, 1996). From the perspective of policy structure, an aggregated policy-structure diagram of the WBI integration model is presented in Figure 1; it demonstrates the conceptual representation of the associations among faculty groups and concern rates.

![Figure 1. A policy structure diagram that shows the associations among faculty concern rates](image)

In the policy-structure diagram, the rectangular sectors represent the structures by which the behaviors of faculty members are determined by several variables. The circular sectors in the diagram represent the structures by which rates of concerns are determined by a number of associated variables. The solid lines represent the material or physical flows that move between the sectors, faculty members. The lines represent information flows transmitted among sectors and the effects of the concern rates. For instance, the Impact Concern Rate is based on the values of the two associated factors, “Advanced training” and “Administrative support for impact.” After the value of the Impact Concern Rate is decided upon, the sector sends information, the “Influence on Impact Concern Rate,” to the “Behaviors of Faculty Implementer” sector. The faculty members in the “Faculty Implementer” group then decide their behaviors based on the information they receive from the other sectors. These potential behaviors include moving forward to become “Faculty Integrators” or falling back to become “Faculty Adopters.” The core assumption is that the more “Advanced training” and “Administrative support for impact” that are provided, the lower the “Impact Concern Rate” will be. As a result, more “Faculty Implementers” are moving forward to become “Faculty Integrators” than are falling back to become “Faculty Adopters.”

Another key concept involved in understanding a complex system is feedback structure, i.e., the causal relationships among variables (Richardson, 1996). There are two kinds of causal loops: positive/reinforcing loops and negative/balancing loops. A change in the value of a particular variable in a positive loop will result in a behavioral change in the system that will eventually strengthen the original change. On the other hand, a change in the value of a particular variable in a negative loop will result in a behavioral change in the system that will eventually counteract or weaken the original change. A causal loop diagram for WBI integration is shown in Figure 2. The arrows in the diagram represent the direction of influence. For instance, an arrow that goes from A to B means that A affects B. In
addition, a positive or negative sign next to an arrow implies that a positive or negative association exists between the two connected variables. When two variables are positively related, they behave in the same manner, meaning that as the value of one variable increases or decreases, the value of the other variable increases or decreases as well.

For example, when the value of “Faculty Entering” increases, the value of “Faculty Unaware” also increases since “Faculty Entering” has a positive effect on “Faculty Unaware.” Similarly, when the value of the “Faculty Entering” decreases, the value of the “Faculty Unaware” decreases as well. As we discussed previously, there are two types of feedback loops used in the model: reinforcing positive loops, which are referred to as (R) loops, and balancing negative loops, which are referred to as (B) loops. Take the R1 loop, for example: when the value of “Faculty Unaware” increases, the values of the “Unaware Faculty Dropping” and the “Dropped Unaware Faculty” variables also increase. An increase in the value of “Dropped Unaware Faculty” results in increased values for both “Returning Unaware Faculty” and “Faculty Entering,” which will eventually lead to a further increase in the value of the “Faculty Unaware.”

In line with the notions of complex social systems and feedback-loop structures that have been proposed, this study considers the adoption of WBI by faculty members as a complex social system that involves various interactions among a number of relevant factors. A system dynamics approach provides both individual faculty members and decision makers/administrators with comprehensive insights into how the indicated relevant factors affect the level of their concerns regarding WBI and as a result influence the behaviors of the faculty members during the WBI adoption process.

A simulation model that tests the impacts of support factors

Structure of the simulation model

A simulation model was constructed to investigate the impacts of the factors that decrease faculty concerns about adopting WBI in higher-education settings. Because simulation is flexible enough to accommodate faculty dropout, constraints and uncertainty in the model, and testing of potential impacts of the parameters in the model, simulation-based methods are employed in this study rather than actual data-collection methods such as surveys and interviews. The STELLA modeling software was chosen to build the WBI adoption process model. STELLA has several advantages over other simulation modeling software, including an easier graphical interface, better classification of variables in the system, easier description of the relationships between variables, an automated process for running computations, and the ability to visualize results with graphical output (Carr-Chellman, Choi, & Hernandez-Serrano,
In this model, all faculty members are classified into one of five faculty groups based on their experience with WBI: “Faculty Aware,” “Faculty Unaware,” “Faculty Adopter,” “Faculty Implementer,” and “Faculty Integrator.” In addition, four main faculty concern rates (Information Concern Rate, Personal Concern Rate, Management Concern Rate, and Impact Concern Rate) are included in the model based on their individual effects on the behaviors of particular faculty groups. Support factors were incorporated into the simulation model at different concern stages during the adoption process. For instance, a Faculty Adopter may experience a management concern as he or she moves toward being a Faculty Implementer, his or her concern rate would be determined based on the combinations of three support factors: the effects of the administrative support for management, the effects of faculty incentives, and the effects of peer tutoring.

User interface of the simulation model

A user-friendly interface was developed for the proposed model that is easy for administrators and faculty members to operate. The interface was designed to help users develop a general understanding of WBI through customized simulations and may also be used to help administrators select appropriate policies to facilitate the use of WBI. The interface, as presented in Figure 4, includes two main parts: the parameter adjustment and the simulation outcome.

The parameter adjustment, which is located at the bottom of Figure 4, includes four divisions that correspond to the different levels of faculty concerns: “Information Concern Factors,” “Personal Concern Factors,” “Management Concern Factors,” and “Impact Concern Factors.” Each concern rate can be changed by adjusting the slide bars that represent the key support factors affecting each concern rate. For example, for the support factor, Basic Training, which falls under “Information Concern Factors,” a user can select a value between 0 and 30, indicating the total number of hours of technical training per semester. The ranges used were determined based on a review of the
literature on the diffusion of online learning and technology integration. Table 2 summarizes the support-factor ranges included in the simulation model. The default value for “Basic Training” is 15 hours per faculty member per semester. If an administrator intends to increase the number of training hours to 20 hours per semester, then the slide bar should be set to 20. The parameter adjustment also includes a brief description of each variable; for example, the “Technology Support” factor, which falls under “Information Support Factors,” includes the information that one technician should be provided for every 100 faculty members. By referring to the brief descriptions provided, a user will understand the meaning of the variable “Technology Support” and make the proper adjustments for his or her simulation. For example, by setting the value of “Technology Support” is set to 5, the user assumes that for every 100 faculty members a total of five technicians are assigned to provide technical support.

![Simulation model interface](image)

**Figure 4.** Simulation model interface

The simulation outcome, which is located at the top of Figure 4, shows the results of the simulation. Users can select values for all the variables based on their own assumptions, and click the “Click to Run” button at the bottom to run the simulation. The model then automatically performs the requested simulation and generates corresponding graphs and numerical information when the simulation is complete. The X-axis represents the time period for the simulation; for example, the X-axis of the graph in Figure 4 which has an index of 0 to 72 indicates that the simulation time frame is 72 months. The Y-axis of the graph represents the values of the five faculty groups at individual points of time. On the top of each of the four sections of concern factors, an information display bar is presented to show the assigned value of the concern rate associated with a particular section in a particular simulation. For instance, the “Information Concern Rate” is 0.5, as shown on the information display bar, when the value of the “Technology Support” is equal to 3 and the value of the “Basic Training” is equal to 15.
Table 2. Ranges for Support Factors

<table>
<thead>
<tr>
<th>Concern Rate</th>
<th>Factor</th>
<th>Range</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Concern Rate</td>
<td>Technology Support</td>
<td>0-20</td>
<td>Technicians for every 100 faculty members</td>
</tr>
<tr>
<td></td>
<td>Basic Training</td>
<td>0-30</td>
<td>Hours of basic technical training</td>
</tr>
<tr>
<td>Personal Concern Rate</td>
<td>Instructional Support</td>
<td>0-20</td>
<td>Instructional specialists for every 100 faculty members</td>
</tr>
<tr>
<td></td>
<td>Intermediated Training</td>
<td>0-45</td>
<td>Hours of training in instructional use of technology</td>
</tr>
<tr>
<td>Management Concern Rate</td>
<td>Administrative Support</td>
<td>0-1</td>
<td>Level of satisfaction with motivational support</td>
</tr>
<tr>
<td></td>
<td>Faculty Incentive</td>
<td>0-3500</td>
<td>Amount of reward per course</td>
</tr>
<tr>
<td></td>
<td>Peer Tutoring</td>
<td>0-30</td>
<td>Hours working with faculty mentors</td>
</tr>
<tr>
<td>Impact Concern Rate</td>
<td>Administrative Support</td>
<td>0-1</td>
<td>Level of satisfaction with assessment, collaboration, and redesign support</td>
</tr>
<tr>
<td></td>
<td>Advanced Training</td>
<td>0-60</td>
<td>Hours of training for refining current use of technology</td>
</tr>
</tbody>
</table>

Examples of the simulation run

The simulation environment for the base run was developed based on a university with a total of 600 full-time faculty members. Since it was difficult to determine the specific number of faculty members in each of the five WBI faculty groups, researchers in this study generally assumed that the initial number of faculty members in each of the five WBI faculty groups was the same: 120. Figure 5A shows the result of the base simulation run, in which we assumed that the university administrators provided the proper support and managed to keep all of the concern rates at 0.5. The X-axis represents the time in months, while the Y-axis represents the number of faculty members in each of the faculty groups. Figure 5A shows that the number of faculty members in the Faculty Unaware (Graph 1) and Faculty Aware (Graph 2) groups decreased noticeably over a time period of 18 months, while the number of faculty members in the Faculty Adopter (Graph 3) and the Faculty Integrator (Graph 5) groups increased considerably within the same period of time.

Figure 5A. Base Run

Figure 5B. Simulation Result—First Scenario
For the simulation, researchers created a couple of possible scenarios. In the first scenario, it is assumed that the university administrators intended to encourage the Faculty Aware individuals, the faculty members who have been aware of WBI, to become Faculty Adopters, that is faculty members who have started to use WBI, within a relatively short period of time. As a result, the university administrators aimed to reduce the concern rate of the individuals in the Faculty Aware group (the personal concern) to 0 by providing the maximum level of “Instructional Support” and “Intermediate Training.” Figure 5B shows the results of the first simulation run. The results indicate that the number of Faculty Adopters (Graph 3) dramatically increased over the first four months. This result implies that more faculty members will adopt WBI when suggested supports such as “Instructional Support” and “Intermediate Training” are provided.

In another scenario, it is assumed that the university administrators were planning to encourage more Faculty Adopters, faculty members who have been starting to apply WBI, to become Faculty Implementers, that is, faculty members who regularly use WBI in their teaching practice, within a relatively short period of time. As a result, the university administrators aimed to reduce the concern rate of the Faculty Adopters (management concern) to 0 by providing them with the maximum levels of the relevant types of support: “Administrative Support for Management,” “Faculty Incentives,” and “Peer Tutoring.” Figure 6B shows the results of this simulation run. The result indicates that the numbers of both Faculty Adopters (Graph 3) and Faculty Implementers (Graph 4) increased within the same period of time. This implies that more “Faculty Adopters” are persuaded to become “Faculty Implementers” when the management concern rate is decreased to 0. The results of the simulation also show that a decrease in the management concern rate smooths the progress of faculty members adopting WBI. For instance, in Figure 6B, the total number of “Faculty Integrators” (Graph 5) jumps to around 180 in the eighteenth month. After the fourteenth month, the majority of the faculty members choose to become “Faculty Adopters,” “Faculty Implementers,” or “Faculty Integrators.”

**Conclusion**

The main purpose of this simulation model was to show administrators and other decision makers how a change in the type of support factors provided to faculty can affect the adoption of WBI. This model will help administrators to better understand the dynamics of the various related to the adoption of WBI by their faculty members. By using this model, administrators can plan how many faculty members can move from one stage to the next within a predetermined time period. These results will provide them with the evidence they need to persuade as many faculty members as possible to become faculty integrators and will convince them to fund workshops and improve incentives to allow new technology to be adopted. Although this simulation model was designed mainly to help administrators of educational institutions make good decisions that will result in increased use of WBI, it is also
expected to be useful for individual faculty members. By using this simulation model, individual faculty members can get an idea of the kinds of support factors available and how much these factors may decrease their concerns. For example, a faculty member with information concerns could identify the best combination of support factors to lower those concerns and request that administrators provide those support factors such as more technology support and basic training opportunities. The model will help researchers and educational administrators to diagnose and evaluate the potential impacts of factors that are predicted to lower faculty concern and contribute to the successful adoption of WBI by the entire faculty.

The simulation model developed for this study suggests several possibilities for future studies. First, a more accurate model that represents the actual change process should be developed, since unknown factors may still remain. More specifically, further efforts such as surveys and interviews with faculty members are needed to clarify the factors that affect each stage of concern. Second, future studies should include relative importance of the support factors in the model. Because the each support factor may have somewhat different significance, future studies should include the relative importance that each factor may have when it is used in the model. Third, interrelations between factors should be thoroughly investigated. It is assumed that each support factor such as faculty incentives and peer tutoring contributed individually to decreases in faculty concerns. However, in practice, these factors may be interrelated and thus, future studies should identify these interrelationships. Therefore, future study should find the interrelationships. Lastly, more factors should be incorporated into the model. Each factor may include various sub-factors, and future studies should include sub-factors that may affect the concerns of the faculty members in order to build a better representative model of what occurs in reality.

References


Hall, E. (1978). *The study of teachers’ concerns and consequent implications for staff development*. Austin, TX: The University of Texas, Research Development Center for Teacher Education.


Identification of Dysfunctional Cooperative Learning Teams Using Taguchi Quality Indexes

Chin-Min Hsiung
Department of Mechanical Engineering, National Pingtung University of Science and Technology, Pingtung, 91207, Taiwan // xiong@mail.npust.edu.tw

ABSTRACT
In this study, dysfunctional cooperative learning teams are identified by comparing the Taguchi “larger-the-better” quality index for the academic achievement of students in a cooperative learning condition with that of students in an individualistic learning condition. In performing the experiments, 42 sophomore mechanical engineering students are randomly assigned to the two learning conditions and are formed into mixed-ability three-member groups. The probability of erroneously identifying a functional team as dysfunctional is quantified using a control chart. The identification results are verified by examining the smaller-the-better and larger-the-better Taguchi quality indexes for the students’ off-task behavior frequency and attitudes toward cooperative learning, respectively. The methodology and findings presented in this study provide a useful basis for the development of technology-assisted solutions for monitoring and analyzing the performance of cooperative learning teams.

Keywords
Academic achievement, Cooperative learning, Dysfunctional team, Taguchi quality index

Introduction
The underlying principle of cooperative learning is one of interdependence (Johnson & Johnson, 1989; Smith, Sheppard, Johnson, & Johnson, 2005). Positive interdependence (i.e. cooperation) prompts an interaction effect in which the individuals within the group encourage one another and facilitate one another’s efforts to learn (Johnson, Johnson, & Smith, 1998; Smith et al., 2005). Many studies have shown that when implemented successfully, cooperative learning fosters a spirit of helping, encouraging, and mutual support, and results in a notable improvement in the learning performance of each team member (Felder, 1995; Felder, Felder, & Dietz, 1998; Johnson, Johnson, Roy, & Zaidman, 1986; Johnson et al., 1998; Prince, 2004; Smith et al., 2005; Springer, Stanne, & Donovan, 1999).

However, when students with different genders, abilities, personalities, learning styles, and work ethics are forced to work together as a group, difficulties invariably arise. In the worst case scenario, some students with an initially positive attitude toward cooperative learning may begin to doubt its effectiveness, while others who were initially resistant to cooperative learning may become even more so (Felder & Brent, 1994). Cooperative teams having such members invariably become dysfunctional over time, leading to an inevitable loss in the benefits of cooperative learning.

As a result, it is desirable for instructors to identify dysfunctional teams at their earliest stages of development such that they can implement appropriate remedial actions in a timely manner. School administrators and education researchers also have an interest in identifying dysfunctional teams. For example, school administrators may wish to analyze the degree of success (or failure) of cooperative learning within their campus in order to tailor their existing teaching strategies or to formulate new strategies for the future. Similarly, in studying the interactions and dynamics within cooperative teams, education researchers may find it useful to identify dysfunctional teams such that their effects can be excluded from the analysis.

Consequently, a requirement exists for a reliable method of identifying dysfunctional cooperative teams to support the developers of educational systems in designing technology-assisted solutions for educators and academics wishing to monitor and analyze the performance of cooperative learning teams.

Purpose of study
Given the importance of identifying dysfunctional teams, the problem of examining how this might actually be achieved in practice has received surprisingly little attention. Some researchers (Hwong, Caswell, Johnson, &
Johnson, 1993) suggested gauging the extent of goal and resource interdependence (i.e. the sharing of a common learning goal and common learning resources) among the individual team members as a means of establishing the success (or otherwise) of the cooperative learning condition. Meanwhile, other researchers (Johnson, Johnson, Stanne & Garibaldi, 1990) analyzed the interaction patterns among the students within the group in order to test for the presence of a successful cooperative learning condition. However, crucially, goal interdependence, resource interdependence, interaction patterns and all the other measures used in previous studies indicate only the success or failure of the cooperative learning condition, i.e. they provide no clues as to the effect of this cooperative learning mode on the academic performance of the individual teams.

Instructors can often identify dysfunctional teams intuitively based upon their informal observations of the group dynamics within the classroom. However, formal measurement and assessment procedures are required to provide a more systematic and objective basis on which to formulate instructional decisions. Kaufman, Felder, & Fuller (2000) asserted that peer ratings are an effective means of identifying dysfunctional cooperative learning teams. Similarly, Haller, Gallagher, Weldon, & Felder (2000) suggested asking students to write “minute papers” to indicate whether or not they felt their teams to be functioning in a healthy and efficient manner. However, the reliability of both methods is fundamentally dependent upon the truthfulness and objectivity of the responses received from the students (Marin-Garcia and Lloret, 2008; Smith et al., 2005).

By contrast, the academic achievement of a group of students can be measured objectively simply by referring to the students’ previous test scores. Academic achievement reflects not only the level of each individual student’s learning performance, but also the degree of cooperation within the group. Yager, Johnson, & Johnson (1985) reported that cooperation within a team promotes a Group-to-Individual knowledge transfer which improves the mean academic performance of the team. In addition, since all of the members in a successful cooperative learning team master the collective knowledge possessed within the team as a result of this knowledge transfer mechanism, it seems reasonable to infer that the test scores within the team will exhibit a greater uniformity than those of students in an individualistic learning condition (Archer-Kath, Johnson, & Johnson, 1994). Thus, the present study asserts that the students’ level of academic achievement, in particular the mean and standard deviation of the students’ test scores, represents a viable means of identifying dysfunctional cooperative teams.

However, while the mean and standard deviation of the test scores within a group provide undoubted insights into the academic performance of the team, both metrics represent only one aspect of the students’ learning performance. Specifically, the mean indicates the central tendency of the test scores, while the standard deviation represents their uniformity. Thus, using one metric alone, or using both metrics together but in an inappropriate way, may easily lead the instructor (or education researcher) to reach false conclusions regarding the degree of cooperation and support within the group. Taguchi and other Quality Engineering researchers have advocated various theories and practices for combining quality characteristics in such a way as to obtain an improved quality measure (Fowlkes & Creveling, 1995; Lochner & Matar, 1990). The resulting quality indexes have been used throughout industry to improve the quality of a wide variety of products and processes (Phadke, 1989; Wu & Wu, 2000). In this study, Taguchi quality indexes are used as a means of identifying dysfunctional teams within an educational context.

Students within a successful cooperative learning team not only exhibit a greater intrinsic motivation to learn than those who study alone (Johnson & Johnson, 1978; Smith et al., 2005), but also attain a higher level of academic achievement (Smith et al., 2005; Yager et al., 1985). However, in dysfunctional cooperative teams, some team members stop cooperating with their peers and work alone, while others simply socialize with members of their own team (or others) and spend significantly less time on-task as a result. In the worst-case scenario, the behavioral problems of one team member may prompt conflicts among the other members, thereby reducing the learning performance of the entire team. As a result, students within a dysfunctional cooperative team tend to achieve a lower academic performance than those who study in an individualistic learning mode. Consequently, in developing a methodology for identifying dysfunctional teams, this study considers both cooperative and individualistic learning conditions and deems a team to be dysfunctional if its members exhibit a lower academic performance than that of students working alone (as evaluated using the Taguchi larger-the-better quality index).

Off-task behavior frequency and attitudes toward cooperative learning

The time that students spend on-task has long been considered an indication of the students’ attitudes toward the subject matter. The positive interdependence goal structure inherent in successful cooperative learning teams
(Johnson & Johnson, 1999), and the resulting actions of helping, encouraging, and supporting one another, create a strong sense of team identity which is instrumental in enhancing accountability at both the individual and the group level (Johnson et al., 1998; Smith et al., 2005). Consequently, compared to students in an individualistic learning environment, students within a successful cooperative learning team tend to exhibit less off-task behavior in class (Hwong et al., 1993). However, many obstacles exist in establishing a successful team, including a lack of team maturity, the history or old norms of the team, the mixed motivations of the team members, obstructive individual behaviors, and so forth (Johnson & Johnson, 2002). Consequently, an increased level of off-task behavior is to be expected in dysfunctional cooperative learning teams. Therefore, in this study, the Taguchi-based identification of dysfunctional cooperative learning teams is validated by reference to the results obtained from off-task behavior frequency observations performed within the classroom.

Previous research has shown that when students feel supported by their peers, cooperative attitudes are more likely to emerge. For example, students who perceive themselves to participate frequently in cooperative learning tasks compared to those who do not sense a greater degree of encouragement from their peers (Johnson, Johnson & Anderson, 1983; Johnson et al., 1986). Therefore, in this study, the students’ perceptions of the level of personal support received from their peers are also used as a means of verifying the results obtained from the Taguchi-based approach for the identification of dysfunctional cooperative learning teams.

### Taguchi quality indexes

Taguchi method is a methodology for finding the optimum settings of the control factors to make the product or process insensitive to noise factors. It has been successfully applied to a wide variety of problems in many different fields, including the electronics, automotive, photography, steel, and chemical processing fields. In fact, the application of the Taguchi method is widely regarded as one of the major factors in accounting for the international dominance of Japanese companies in these industries and others (Phadke, 1989).

Taguchi proposed various generic signal-to-noise (SN) ratios for evaluating the quality of engineering products or processes. Of these ratios, those most commonly used include the smaller-the-better (SB) ratio, the larger-the-better (LB) ratio and the nominal-the-best (NB) ratio (Fowlkes & Creveling, 1995). The NB ratio takes into account the process mean and variance and provides an easy-to-use performance criterion (Lochner & Matar, 1990; Wu & Wu, 2000). In nominal-the-best type problems, the quality characteristic is continuous and nonnegative. Its target value is nonzero and finite. This type of problem occurs frequently in engineering design. Keeping the water in a tank at a constant temperature and achieving target thickness in an electroplating job are two typical nominal-the-best type problems. The objective function (the NB ratio) to be maximized for such problem is written as $SN = 10 \log (\mu^2 / \sigma^2)$, where $\log (\cdot)$ is the base 10 logarithm function, $\mu^2$ is the square of the mean of the response and $\sigma^2$ is the variance of the response. It is called the signal-to-noise ratio because $\sigma^2$ is the effect of noise factors and $\mu^2$ is the desirable part of the data (the signal). Maximizing SN is equivalent to minimizing sensitivity to noise factors. Moreover, the NB ratio provides the basis on which the other two SN ratios are derived (Lochner & Matar, 1990; Wu & Wu, 2000).

In detecting the presence of dysfunctional cooperative learning teams, this study utilizes both the SB ratio and the LB ratio. In smaller-the-better type problems, the response values or quality characteristics have non-negative values and the desired value of the response is zero (Lochner & Matar, 1990). In the present study, the students’ off-task behavior frequency falls neatly into this category of problem. The smaller-the-better SN ratio is formulated as $SN = -10 \log (\Sigma y_i^2 / n)$, where $\Sigma$ is the summation operation, $n$ is the number of data points, and $y_i$ is the $i$-th individual response value. The SB ratio can also be written as $SN = -10 \log (\sigma^2 + \mu^2)$. In other words, the SB signal-to-noise ratio takes account of both the variability in the response data and the closeness of the average response to the target value (i.e. zero). Specifically, the SN ratio increases as the variability and average value of the response reduces (Fowlkes & Creveling, 1995; Lochner & Matar, 1990). For larger-the-better type problems, the response values or quality characteristics have non-negative values and the aim is to increase the value of the response toward the largest number possible (preferably infinity). In the present study, this definition neatly fits the problem of evaluating the students’ academic achievement and their attitudes toward cooperative learning. Larger-the-better problems are the reciprocal of smaller-the-better problems, and thus the larger-the-better SN ratio is formulated as $SN = -10 \log (\Sigma y_i / n)$ (Fowlkes & Creveling, 1995; Lochner & Matar, 1990). Note that regardless of the SN ratio chosen as a quality index, the objective of the Taguchi design procedure is always to increase the value of the SN ratio, i.e. a
higher SN ratio is always indicative of a better quality (Fowlkes & Creveling, 1995; Lochner & Matar, 1990; Phadke, 1989; Wu & Wu, 2000).

Method

Subjects

The experimental stage of this study was conducted in the spring semester of 2006 using 42 mechanical engineering sophomore students from National Pingtung University of Science and Technology in southern Taiwan. The students were randomly assigned to a cooperative learning condition or an individualist learning condition such that each condition contained 21 students. Within the cooperative learning condition, the 21 individuals were assigned to three equally-sized “high-ability”, “medium-ability” or “low-ability” groups in accordance with their average test scores in the previous semester. Within each ability group, one member was chosen at random and assigned to a cooperative learning team. This process was repeated iteratively until each of the 21 students had been assigned to a learning team. Thus, on completion of the assignment process, the 21 individuals were organized into a total of seven learning teams, with each team comprising one high-ability member, one medium-ability member and one low-ability member. Having been assigned to one of the seven groups, the students were informed that they were to remain within the same group for the entire semester and were to work in a cooperative fashion in solving the learning tasks assigned to them. For ease of comparison, the 21 students in an individualistic learning condition were also randomly assigned to seven “teams” stratified in terms of the members’ academic ability. However, in this case, the students were instructed to work alone at all times. The results presented in Table 1 confirm the heterogeneous nature of the cooperative and individualistic learning teams. In addition, the results obtained from a t test showed no significant differences in the mean test scores of the students within the two conditions, t = 0.15, p = 0.88. Therefore, it was inferred that the two learning conditions were equivalent in terms of the students’ academic abilities prior to the experimental stage of the study.

<table>
<thead>
<tr>
<th>Team</th>
<th>Cooperative Learning</th>
<th>Individualistic Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ability group</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>A</td>
<td>75</td>
<td>70</td>
</tr>
<tr>
<td>B</td>
<td>73</td>
<td>67</td>
</tr>
<tr>
<td>C</td>
<td>71</td>
<td>68</td>
</tr>
<tr>
<td>D</td>
<td>77</td>
<td>71</td>
</tr>
<tr>
<td>E</td>
<td>76</td>
<td>71</td>
</tr>
<tr>
<td>F</td>
<td>82</td>
<td>69</td>
</tr>
<tr>
<td>G</td>
<td>77</td>
<td>68</td>
</tr>
</tbody>
</table>

Course

The experimental study was confined to the Planar Dynamics course, a mandatory course for each of the 42 students. The course comprised four teaching units, namely Unit 1 - Kinematics, Unit 2 - Energy Method, Unit 3 - Force and Acceleration Method, and Unit 4 - Impulse and Momentum Method. Units 1 and 3 each had a duration of six weeks, while Units 2 and 4 both lasted for three weeks. In addition to the regular 3-hr/wk day-time classes, the students also attended mandatory homework classes (7-hr/wk) in which they studied in accordance with their assigned learning condition under the supervision of two teaching assistants.

Independent variables

In this study, the independent variable is the learning method (i.e. cooperative or individualistic). In the cooperative condition, the students were instructed to work together as a team, thereby promoting that each team member worked diligently, mastered the assigned material, freely offered their ideas and suggestions, listened to one another, shared
ideas and materials, and praised and helped one another. The cooperative learning teams were carefully implemented in such a way as to develop positive interdependence, face-to-face interaction, individual accountability, social skills, and group processing (Johnson & Johnson, 1978, 1989, 1999). Note that a detailed review of the implementation procedures required for successful cooperative learning teams can be found in previous studies (Oakley, Felder, Brent, & Elhajj, 2004; Smith et al., 2005), and thus the details are deliberately omitted here.

In the individualistic learning condition, the students worked on their own, avoided interaction with their peers, sought help only from the instructor or teaching assistants, and worked at a self-regulated pace. Interaction between the students in the individualistic learning condition was actively discouraged, and the students were constantly reminded to work alone when performing the assignments given to them in class or in the evening homework sessions.

Throughout the entire semester, only one student dropped out of the course. In addition, the attendance records showed that the two groups of students in the different learning conditions had an approximately equal attendance rate, and thus an equal learning time. Furthermore, for the duration of the course, two trained teaching assistants were engaged to help the instructor enforce the two learning conditions. Specifically, the assistants moved around the class encouraging students in the cooperative condition to work together in solving the assigned tasks, while reminding those in the individualistic condition to study alone.

**Dependent variables**

The dependent variables in the present study include the students’ academic achievement, their off-task behavior frequency, and their attitudes toward cooperative learning. The students’ academic achievement was measured by means of four unit tests, given at the end of each teaching unit, respectively. Each test lasted for three hours and consisted of six to eight engineering problems chosen from the exercise section of the course textbook (Hibbeler, 2004). In completing each test, the students in each learning condition worked individually and were assigned a score in accordance with their own performance. A solution flowchart was prepared in advance by the instructor for each test item to indicate the key intermediate stages and results in the solution procedure. Using these flowcharts, the students were awarded partial credits even if their final solutions were incorrect. The test papers were graded independently by the two teaching assistants, and the two sets of results were then compared to ensure their reliability. In analyzing the test results obtained over the course of the semester, the scores achieved by each student for the four unit tests were averaged to obtain a single measure of the student’s academic achievement.

The frequency of the students’ off-task behavior was measured by conducting classroom observations (Hwong et al., 1993; Marin-Garcia and Lloret, 2008; Slavin, 1992) during the homework classes. The observation sessions were conducted on five separate occasions around the mid-point of the semester. In performing the observations, off-task behavior was defined as one or more students within a cooperative team talking with members of another team, walking around the classroom, breaking eye contact with their team members or the learning materials, dozing, or simply socializing. The observation sessions were conducted by the two teaching assistants using an off-task record form and an MP3 device with a 20-second cued loop designed to provide a consistent aural cue for a 10-second “observe” period and a 10-second “record” period. Inter-observer reliability was calculated by dividing the total number of agreements between the results obtained by the two observers by the sum of the agreements and disagreements (Hwong et al., 1993; Slavin, 1992). The average reliability index was found to be 0.85, and thus it was concluded that the observation results were sufficiently consistent for analysis purposes (Slavin, 1992). In analyzing the observation results, a point was given for each off-task behavior observed, and the total number of points recorded by the two observers for each team was then averaged to obtain an overall result. Finally, the results obtained for each team over the five observation sessions were used to calculate the corresponding value of the smaller-the-better SN ratio in accordance with the formula given in Section Taguchi-quality-indexes.

The attitudes of the students toward cooperative learning were evaluated at the end of each teaching unit using the peer-personal-support subscale developed by Johnson et al. (1983). The subscale comprised five items (Cronbach alpha reliability: 0.78) designed to determine each respondent’s perceptions regarding whether or not the other team members cared about their feelings, liked them the way they were, really cared about them, liked them as much as they liked the other members in the team, and thought it was important to be their friend, respectively. Each item on the subscale was configured in the form of an affirmative statement and was evaluated using a 5-point scale, where 1
indicated the respondent’s absolute disagreement with the statement and 5 indicated the respondent’s absolute agreement. Thus, a higher score revealed a more favorable attitude on the part of the respondent to the cooperative learning approach. In analyzing the survey results for each student, the scores obtained over all of the items were summed and the result was then divided by the total number of items on the subscale (i.e. 5). The results obtained for each respondent over the four surveys were then averaged to obtain a single measure of the respondent’s attitude toward cooperative learning. Finally, the results obtained for each team over the four surveys were used to calculate the corresponding value of the larger-the-better SN ratio in accordance with the formula given in Section Taguchi-quality-indexes.

Results and Discussion

Traditional method

Table 2(a) presents a statistical analysis of the unit test scores for the seven teams in the cooperative learning environment. Note that in compiling the data presented in this table, the mean score for each team was obtained by averaging the scores for the three team members (N=3) within the group, while the standard deviation was obtained by taking the square root of the variance of the scores obtained by each team member in the group. Table 2(b) presents the statistics of the test scores of the 21 students in the individualistic learning condition.

Table 2(a). Statistical analysis of the scores for the seven cooperative teams

<table>
<thead>
<tr>
<th>Cooperative Team</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>48.92</td>
<td>13.76</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>61.75</td>
<td>5.94</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>62.50</td>
<td>13.21</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>51.50</td>
<td>13.05</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>66.00</td>
<td>9.12</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>68.92</td>
<td>8.38</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>66.50</td>
<td>6.95</td>
</tr>
</tbody>
</table>

Table 2(b). Statistical analysis of the scores for the individualistic learning condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individualistic Learning</td>
<td>21</td>
<td>59.42</td>
<td>8.73</td>
</tr>
</tbody>
</table>

Conventional wisdom suggests that successful collaboration within a cooperative learning team not only improves the mean academic achievement of the team, but also enhances the uniformity of the mean achievement of the individual team members. Thus, the results presented in Table 2(a) suggest that cooperative teams A and D are the most likely candidates for dysfunctional cooperative learning teams since they not only have a lower mean score than the average score of the 21 students in the individualistic learning condition (i.e. 48.92 and 51.50 for teams A and D, but 59.42 in the individualistic learning condition), but also have a higher standard deviation (i.e. 13.76 and 13.05 for teams A and D, but 8.73 for the students in the individualistic learning condition). Teams C and E are the next most likely candidates for dysfunctional teams since even though they have a relatively high mean score (i.e. 62.5 and 66.0, respectively), they both have a large standard deviation (i.e. 13.21 and 9.12, respectively). In general, however, it is difficult to reach a definitive conclusion regarding the status of a team if the team has only a small mean score or a large standard deviation (e.g. teams C and E).

Taguchi method

Table 3(a) presents the Taguchi SN ratio values for the three dependent variables in this study. Note that the results are obtained by substituting the values of the dependent variables for each team into the formulae given in Section Taguchi-quality-indexes. Note also that the value of n given at the head of each column indicates the number of data points used to compute the value of the SN ratio for the corresponding dependent variable. Thus, for the off-task frequency variable (see the third column in Table 3(a)), n has a value of n=5 since the off-task behavior was evaluated over five observation sessions. Meanwhile, the SN ratios for the academic achievement and perceived peer support variables were computed using n=3 since each team comprises three team members and each member contributes one data point to the SN calculation. Table 3(b) presents the SN values of the basis of comparison. Note
that since the off-task behavior frequency was observed only for those individuals within a cooperative learning team, the data presented in the third column of this table is actually the mean SN value of the off-task behavior frequency of the seven cooperative learning teams. As a result, the corresponding SN ratio is computed using $n=35$, i.e. five observations for each of the seven teams. By contrast, the SN ratios for the academic performance and perceived peer support, respectively, are both computed using $n=21$, corresponding to the 21 students within the individualist learning condition. The results presented in the second column in Table 3(a) (i.e. the SN ratio for academic achievement) suggest that cooperative learning teams A and D are dysfunctional since their SN values (32.77 and 33.48, respectively) are smaller than the mean SN ratio (35.15) of the students in the individualistic learning condition.

Table 3(a). SN ratio for dependent variables of cooperative teams

<table>
<thead>
<tr>
<th>Cooperative Team</th>
<th>Academic achievement (n=3)</th>
<th>Off-task frequency (n=5)</th>
<th>Peer personal support (n=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>32.77</td>
<td>-14.44</td>
<td>10.85</td>
</tr>
<tr>
<td>B</td>
<td>35.70</td>
<td>-11.40</td>
<td>14.42</td>
</tr>
<tr>
<td>C</td>
<td>35.20</td>
<td>-2.04</td>
<td>11.56</td>
</tr>
<tr>
<td>D</td>
<td>33.48</td>
<td>-15.19</td>
<td>10.97</td>
</tr>
<tr>
<td>E</td>
<td>36.11</td>
<td>-2.55</td>
<td>10.76</td>
</tr>
<tr>
<td>F</td>
<td>36.58</td>
<td>-10.00</td>
<td>12.56</td>
</tr>
<tr>
<td>G</td>
<td>36.32</td>
<td>0.97</td>
<td>11.60</td>
</tr>
</tbody>
</table>

Table 3(b). SN ratio for dependent variables of the comparison basis

<table>
<thead>
<tr>
<th>Condition</th>
<th>Academic achievement (n=21)</th>
<th>Off-task frequency (n=35)</th>
<th>Peer personal support (n=21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individualistic Learning</td>
<td>35.15</td>
<td>-11.03</td>
<td>10.74</td>
</tr>
</tbody>
</table>

Verification: off-task behavior frequency

Off-task or on-task behavior frequency observations have been used in many previous studies to investigate the impact of cooperative learning (Johnson, Johnson, Ortiz, & Stanne, 1991; Archer-Kath et al., 1994). In general, the results have shown that successful cooperative learning teams exhibit less off-task behavior in class (Hwong et al., 1993). However, these studies did not prescribe any formal rules to enable the reliable identification of dysfunctional teams. In this study, it is argued that dysfunctional teams exhibit a greater amount of off-task behavior than successful cooperative teams, and consequently have a lower SN ratio. Thus, the results presented in Table 3(a) support the earlier findings that cooperative teams A and D are dysfunctional since their off-task frequency SN ratios (i.e. -14.44 and -15.19, respectively) are much (3 dB) lower than those of the other teams.

In this study, the observation sessions were deliberately performed around the middle of the semester in order to allow the cooperative teams sufficient time to mature (Ortiz, Johnson, & Johnson, 1996). However, they could actually have been conducted at any time during the semester. In other words, the instructor is free to choose both when and how often to conduct the observation sessions. In addition, the observation procedure is straightforward and systematic. Each observation procedure takes only a minute to complete for one team, and several minutes at most to complete for the entire class. Moreover, the observation procedure does not depend on the students’ cooperation and has scarcely any impact on the normal teaching activities within the class. These advantages render off-task behavior frequency observations a reliable and convenient tool for the provision of supporting evidence when attempting to identify dysfunctional teams.

Verification: attitudes toward cooperative learning

Previous research has shown that students who participate frequently in cooperative learning tasks sense a greater amount of encouragement and academic help from their peers (Johnson et al., 1983). Therefore, in this study, it is argued that members of a dysfunctional team have a less favorable perception of the cooperative learning approach than those of a functional team. Thus, the peer support data presented in the fourth column of Table 3(a) confirm the earlier findings that cooperative teams A and D are dysfunctional.
It is observed in Table 3(a) that team B also has a relatively low off-task behavior frequency SN ratio (i.e. -11.40). However, it has a larger than average SN ratio for academic achievement (i.e. 35.70 > 35.15) and a relatively positive attitude toward cooperative learning (i.e. 10.76), but has a larger than average SN ratio for academic achievement (i.e. 36.11 > 35.15). As a result, neither team is considered to be dysfunctional.

Compared with the peer-rating method, attitude surveys such as those conducted in this study are likely to be more favorably received by the team members since they measure each member’s own feelings and emotions rather than asking them to evaluate the behavior and attitudes of their team mates. As a result, the respondents are more likely to answer in an open and truthful manner. However, as with peer-rating methods, the results obtained from an attitude survey are also subjective to a certain extent since different respondents may apply different criteria when interpreting and responding to the survey items. Consequently, it is suggested that while attitude surveys can complement or reinforce the identification results obtained using analytical quality indexes such as the Taguchi SN ratio proposed in this study, they should not be used as the sole basis for identifying dysfunctional teams.

**Absolute quality information**

The SN ratios described in the preceding sections enable a comparison to be made between the “quality” of each team (Fowlkes & Creveling, 1995), e.g. teams A and D perform less successfully than the other cooperative teams. However, to answer the questions “just how much worse are teams A and D than the other teams” or “are teams A and D truly dysfunctional” some form of absolute quality information is required.

Control charts have long been used in the statistical process control field to evaluate the probability of a data point being wrongly classified as failing to satisfy the control limits (Montgomery, Runger, & Hubele, 2001). In a sense, this probability information can be taken as an indication of the absolute quality of each data point (i.e. each cooperative team in the current context). In a control chart, the upper and lower control limits (i.e. UCL and LCL, respectively) and the center line (CL) are defined as follows: $UCL = \mu + k*\sigma/sqrt(n)$, $LCL = \mu - k*\sigma/sqrt(n)$, and $CL = \mu$, respectively, where $\mu$ denotes the population mean, $\sigma$ denotes the population standard deviation, $sqrt()$ denotes a square root operation, $n$ is the sample size, and $k$ is the “distance” of the control limits from the center line and is expressed in multiples of the standard deviation. For the case where the process is in control, the probability of a plotted point falling below the lower limit is equal to 0.02 for $k=2$ and 0.16 for $k=1$.

Table 4(a) indicates the composite academic achievement SN ratios of the seven teams in the cooperative learning condition. Note that in compiling this table, the test scores of the three team members for each unit test are collected and the corresponding value of the SN metric is computed using the larger-the-better formula presented in Section Taguchi-quality-indexes. The composite SN ratio for the team is then obtained by averaging the SN values over the four unit tests. Table 4(b) presents a statistical analysis of the academic performance SN ratio for the 21 individuals in the individualistic learning condition. Since the individualistic condition contains a total of seven notional “teams” and each team has four SN values (i.e. one SN value per unit test), the total number of SN values is equal to $N=28$ (see second column in Table 4(b)).

<table>
<thead>
<tr>
<th>Cooperative Team</th>
<th>N</th>
<th>Composite SN</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>32.47</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>35.10</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>34.80</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>32.41</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>34.62</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>36.20</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>35.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>LCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individualistic</td>
<td>28</td>
<td>34.43</td>
<td>3.30</td>
<td>32.78</td>
</tr>
</tbody>
</table>
In the present study, the academic achievement of the students in an individualistic learning condition is taken as the basis of comparison for identifying dysfunctional cooperative learning teams. As a result, the population mean SN value used in constructing the control chart is given by the average SN value presented in the third column of Table 4(b) while the population standard deviation of the SN values is equal to the standard deviation value given in the fourth column of Table 4(b). In other words, the population mean of the SN ratio is equal to $\mu=34.43$ while the population standard deviation of the SN ratio is equal to $\sigma=3.30$. Note that although the four unit tests were conducted sequentially over the course of the semester, all of the SN values were actually calculated and analyzed at the end of the semester. Notionally, this is equivalent to collecting four SN values for each team at the end of the semester. Thus, the sample size is equal to $n=4$. Consequently, the LCL has a value of 32.78 for $k = 1$ (see fifth column in Table 4(b)). Therefore, the results presented in the third column of Table 4(a) indicate that cooperative teams A and D are truly dysfunctional since their composite SN values, i.e. 32.47 and 32.41, respectively, are less than the LCL. Note that the probability of misclassifying teams A and D is equal to 0.16. In Table 4(a), cooperative team E has the third lowest SN value, i.e. 34.62. The probability of erroneously classifying team E as dysfunctional due to its small SN value is equal to 0.55. Thus, it follows that only cooperative teams A and D are truly dysfunctional.

Research limitations

Generalization of the results of this study is limited by the cultural context of the study (namely a science and technology university in Taiwan), the sample size, characteristics of the subjects, types of tasks, skills of implementing cooperative learning of the instructor, and specific operationalizations of the independent and dependent variables. Nonetheless, despite these limitations, the results obtained in this study are reliable because the following important procedures have been employed: the stratified random assignment of students to conditions, the use of the same instructor to teach both conditions, the use of measurable dependent variables, and the confirmation of the implementation of conditions. In a future study, further investigations will be performed to explore whether the present findings are equally applicable to a broader educational context.

Conclusions

This study has proposed a new method for the identification of dysfunctional cooperative learning teams based on students’ academic achievement utilizing the well-known smaller-the-better and larger-the-better Taguchi quality indexes. The feasibility of the proposed approach has been confirmed by performing a series of experimental investigations using 42 sophomore engineering students. The students’ academic achievement was measured by means of four unit tests, given at the end of each teaching unit, respectively. Taken with the supporting evidence provided by off-task behavior frequency observations and peer support surveys, the larger-the-better Taguchi quality index provides a reliable means of identifying dysfunctional cooperative learning teams. Moreover, the probability of erroneously classifying a functional cooperative learning group as dysfunctional can be estimated using a control chart in which the Taguchi quality index is taken as the quality characteristic.

The method proposed in this study for identifying and verifying dysfunctional teams within a cooperative learning environment can be easily incorporated into a technology-assisted solution for instructors wishing to monitor and analyze the performance of cooperative teams on a continuous basis throughout the course of a semester. Such a system could also assist school administrators in computing the means of the unit tests given at the end of each teaching unit such that any dysfunctional teams can be automatically identified at the end of the semester. Utilizing this information, the administrators could then analyze the degree of success (or failure) of the various cooperative learning initiatives on their campus and modify their existing teaching strategies or formulate new strategies as required. Similarly, in identifying the true nature of the interactions and dynamics within cooperative teams, education researchers would also benefit from an automated solution for the reliable identification of dysfunctional teams such that the effects of these teams could be excluded from the analysis. In practice, the methodology presented in this study is well suited to meeting the requirements of all three groups of individuals in the cooperative learning environment.
Thus, the methodology and findings presented in this study represent a useful basis for educational system developers seeking to design technology-assisted solutions for supporting educators in monitoring and analyzing the performance of cooperative learning teams.

Acknowledgements

This study was supported by the National Science Council of Taiwan under Grants NSC94-2516-S-020-006 and NSC95-2516-S-020-003. The author is also grateful to Director T. H. Wang of NSC for his encouragement in writing this manuscript.

References


Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turing student groups into effective teams. *Journal of Student Centered Learning, 2*(1), 9-34.


Web-based Spatial Training Using Handheld Touch Screen Devices

Norena Martin-Dorta, Jose Luis Saorin and Manuel Contero1
Dpto. de Expresión Gráfica en Arquitectura e Ingeniería, Universidad de La Laguna, Tenerife, Spain // 1Instituto de Bioingeniería y Tecnología orientada al Ser Humano, Universidad Politécnica de Valencia, Valencia, Spain // nmartin@ull.es // jlsaorin@ull.es // mcontero@labdesign.13bh.es

ABSTRACT
This paper attempts to harness the opportunities for mobility and the new user interfaces that handheld touch screen devices offer, in a non-formal learning context, with a view to developing spatial ability. This research has addressed two objectives: first, analyzing the effects that training can have on spatial visualisation using the educational content developed for this pilot study; and second, evaluating the experience of users in the use of handheld touch screen devices and their degree of satisfaction with the on-line course proposed. Two study groups were used: an experimental group, which undergoes a one week training programme, and a control group, which does no spatial training tasks during this period. The results show a significant improvement in average spatial visualisation scores among the experimental group in comparison with the control group. Students value positively the course accomplished and they expressed their preference for these multimedia contents over the conventional pencil and paper formats, and for the on-line learning over a face to face course. They also consider that having materials of this kind available in their study programme subjects and the possibility of accessing this material at any time and in any place to be a positive aspect.

Keywords
Spatial abilities, Spatial visualisation, Mobile learning

Introduction
Progress made in technology, user interfaces and recent developments in the field of communications (Wi-Fi, GPRS, 3G, etc.), have created a wide range of possibilities for users. Devices such as PDAs, mobile phones, videogames consoles, MP3/MP4 or ultra mobile PCs (UMPC), now increasingly incorporate one or more of these technologies available. Furthermore, improvement in screens and new interfaces, together with multimedia and storage capacities for podcasts, videos, photos, files, etc., eases interaction with a huge range of contents (Guerrero, 2006; Johnson, Levine & Smith, 2008).

Play, in its diverse forms, constitutes an important part of children’s cognitive and social development (Piaget, 1951; Provost, 1990). Several authors have analyzed the impact of games on education and there is wide empirical evidence supporting the positive effects of computer games as instructional tools (Amory, Naicker, Vincent & Adams, 1999). Recent studies have brought attention to the educational potential of handheld devices. Some tools have been tested on these devices, indicating they strengthen and support learning in fields such as languages (Thornton & Houser, 2005; Lu, 2008; Chen & Chung, 2008), science and natural history (Facer, Joiner, D., Reid, Hull & Kirk, 2004; Sánchez, Salinas & Sáenz, 2007), and also provide an additional tool in common learning (Corlett, Sharples, Bull & Chan, 2005; Clough, Jones, McAndrew & Scanlon, 2007; Hoff, Wehling & Rothkugel, 2008; Chris, 2008).

Spatial skills may be associated with success in scientific areas (Smith, 1964). Non-academic activities, such as playing with construction toys as a young child and playing three dimensional computer games seem to have strong relationship with spatial visualization ability. The potential of video games or computer games for improving spatial skills have been analyzed by numerous research (Deno, 1995; Sorby, 2007). Most recent research in the field of spatial abilities focuses on how these relate to new technologies (Rafi, Samsudin & Ismail, 2006; Rafi, Samsudin & Said, 2008.; Martin-Dorta, Saorin & Contero, 2008; Rafi & Samsudin, 2009).

Our work focuses on mobility and new user interfaces that handheld touch screen devices offer, in a non-formal learning context, targeting development of spatial ability. This research has addressed two main objectives: first, analyzing the effects that training may have on spatial visualisation using the educational content developed for this pilot study; and second, evaluating students experience while using handheld touch screen devices and their degree of satisfaction with this online learning course proposed.
Spatial abilities and improvement tools

Over the last half century, spatial abilities have been given increasing recognition and, despite the fact that not so much attention has been paid to them as to verbal and numeric abilities, research accentuates their importance in the traditional fields of engineering, technology and art, as well as in almost any other aspect of life. As it has repercussions in almost all scientific and technical fields, spatial abilities remain an active field of study, especially in the engineering area. Considered as a component of intelligence throughout history, we can understand it as the ability of manipulating objects and their parts mentally in a two-dimensional and three-dimensional space. From the point of view of measuring it, we can define it as the ability to imagine rotations of 2D and 3D objects as a whole body (Spatial Relations) and the ability to imagine rotations of objects or their parts in 3D spatial by folding and unfolding (Spatial Visualisation) (Saorín, 2006).

Work done in recent years has been undertaken in the areas listed below:

- Measuring the spatial abilities of students who come into university, by studying the prior educational factors that could have an impact on these results: gender, age, previous experiences, etc.
- Measuring the effect of the Engineering Graphics subjects on the University population.
- The development of multimedia or web-based tools.
- Creating remedial courses with a view to improving the level of students coming into university. This is because it has been found that some students arrive with deficits in this area and that it could be an indicator of possible failure in engineering studies.

Some authors have based their work on the hypothesis that spatial abilities can be improved if the right tools are used; i.e. ones that facilitate an understanding of the concepts and the relations between two and three-dimensional representations. The appearance of new technologies has meant that, since the mid nineties, several different research groups have suggested novel tools for improving spatial abilities.

Because of the breadth of their research and their continuity, it is worth highlighting the work led by Sheryl A. Sorby, in the Michigan Technological University (MTU). As a result of over ten years of research, she has published a ten module manual to improve 3D spatial visualisation (Sorby, Wysocky, & Baartmans, 2003). This allows students to work with isometric exercises by building standardised orthogonal views, cross sections, rotations, etc. with blocks. Apart from the text book, this publication comes with a flash based technology CD-ROM to complement the hard copy. This application offers interactive effects, creating the experience of a computer game for students (Figure 1a).

At the University of Pennsylvania, Hollyday-Darr, Blasko & Dwyer (2000) created “Visualisation Assessment and Training Program” (Figure 1b and Figure 1c). This tool allows us to evaluate three components of spatial abilities: mental rotation, spatial visualisation and spatial perception. Rafi, Samsudin & Ismail (2006) tested the efficiency of this material in a comparative study in 2006. Some years later, these authors also developed and validated a new tool for improving mental rotation (Rafi & Samsudin, 2009).

Professor Stephen W. Crown, from the Pan American University of Texas, hosts a multimedia web site on his university web site, which is comprised of a set of interactive games, tutorial videos and theory lessons that reinforce important concepts of Engineering Graphics. Tests and satisfaction surveys demonstrate that the impact of his tool on students has been a positive one (Crown, 2001) (Figure 1d).

Conolly & Maicher (2005) from Purdue University, developed “Multiview Drawing”, an interactive tutorial for beginners in the world of orthographic view drawing. Given an isometric view of a piece, the programme enables students to draw a view and automatically correct the final result (Figure 1e).

Illuminations (National Council of Teachers of Mathematics, 2000) is a website that was created by the United States National Council of Teachers of Mathematics and the Marco Polo Foundation. The site includes the “Isometric Drawing Tool”, an interactive applet that allows you to dynamically create isometric drawings on a template of dots (Figure 1f).
Figure 1. (a) Sorby’s training manual; (b) and (c) University of Pennsylvania training programme; (d) S. Crown’s interactive game; (e) Multi-view Drawing by Purdue University; (f) Isometric Drawing Tool

In Spain, the University of Burgos offers a Spatial Visualisation Workshop (Department of Engineering Graphics, University of Burgos, 2007), a web-based course made up of five modules: identification of surfaces, identification of views, discrimination, volumes and rotations, counts and a final level devoted to assessment (Figure 2a). At the University of Oviedo, Moran, Rubio, Gallego, Suarez & Martin (2008) have created a set of applications to improve students’ spatial perception using three flash-based web applications: view table, views of a cube and the development of a tetrahedron (Figure 2b).

Since 2004, the Dehaes Research Group (2009) has developed several lines of research focusing on the study of spatial abilities in engineering students. First of all, they conducted a study in the form of short remedial courses, with the main result of improving the spatial abilities of students when they first go to university and to provide a test bench for designing action strategies (Contero, Naya, Company, Saorín & Conesa, 2005; Contero, Company, Saorín & Naya, 2006). Three 6-hour remedial courses were held for students with problems related to spatial abilities, each of which was held in three 2-hour sessions. A sketch-based modelling application called e-CIGRO was used on one of the courses (see Figure 3), developed by the Regeo Research Group (2009). A second experience was carried out during the 2006-2007 academic year, this time based on the use of 3D modelling software (Martin-Dorta, Saorín, & Contero, 2008) (see Figure 4). During the 2008-2009 academic year, we have worked on developing a course that
uses Augmented Reality Technology, exploring a new branch of interfaces in which real elements co-exist with virtual ones. This is an innovative tool that we have used to create an “Augmented Book”, which is a training manual where users can see virtual objects, manipulating them so they can be seen from any possible angle (see Figure 5).

The most recent research in the field of spatial abilities focuses on how these relate to new technologies. The rapid growth of information and communication technologies and the skill that modern students have with these technologies provide new forms of education. The use of small mobile devices (smartphones, videogames consoles, etc.), it’s currently creating huge expectation so many interesting business initiatives and research projects are working on this field (The International Association for Mobile Learning, 2009).

This paper focuses on making the most of opportunities created by mobility and new user interfaces offered by handheld touch screens, in a non-formal, online learning context, concentrating on developing spatial visualisation and attaining the learning objectives promoted by the new European Higher Education Area.
Pilot study

This pilot study has been designed for obtaining experience in order to launch a short remedial course to improve spatial abilities in first year engineering students with underdeveloped abilities at the beginning of their first term at university. The general objective is that, by the end of the course, participant students can achieve a minimum level in their spatial abilities, which should contribute to a successful participation in the regular engineering graphics course, taught during the first term in the majority of undergraduate engineering programmes. The specific objectives of this research have addressed the following aims:

- Analizing the effects that training based on the teaching material described in this work can have on spatial visualisation,
- Assessing the experience of users with handheld touch screen devices, a new format in the area of spatial training, and
- Rating the satisfaction of users with the on-line learning course (mobile learning).

Participants

A group of sixty eight volunteer first year students on the Civil, Electronic and Mechanical Engineering Degree Courses at the University of La Laguna in Spain. Twenty three females and forty five males, mean age 20 (range 17-28). The participants have no previous experience with handheld touch screen devices. This study uses two working groups: an experimental group and a control group. The experimental group consists of 38 students (26 men and 12 women) who have volunteered to do the training programme, and the control group, made up of 30 students (19 men and 11 women), who receive no training during the week that this study lasts.

Materials

Didactic Content Developed for Spatial Training Course

Based on the material published by different researchers in this field, which is summarised in section two of this paper, we have designed a classification of the different types of exercises and we propose a series of modules and levels, in accordance with the difficulty of the tasks to be conducted. Classification and typologies of exercises used in this paper have been taken from several scientific papers (from Sorby and others) which provide a reliable source of validated sets of exercises in spatial abilities research. This is the first time that these set of exercises is used with mobile devices. Didactic content has been optimized for a 3.5 inch touch screen devices.

The course is structured in five modules with several levels each, as shown in Table 1. The first module “Building with blocks”, is composed of block model exercises. The students have to identify the number of blocks that match the one indicated in a drawing (level 1), choosing from four models the one that matches a given numeric coded plan (level 2) or distinguish whether the proposed figures are possible to build in the real world or not (level 3). “Identification of Sides and Views”, the second module of this teaching material, contains exercises in which the students have to identify surfaces and orthogonal views when they have the axonometric view of an object. The third module, “Object Discrimination”, consists of three levels in which the students must try to relate isometric sketches of the object to the right orthogonal views. The fourth module, “Rotations”, addresses mental rotations with exercises in which the students have to identify the orthogonal views or isometric sketches that have been rotated through 90, 180 or 270 degrees from a given axis. The fifth and final module, “Cross Sections”, requires students to identify cross sections made of the objects shown. Each level has an explanatory video that reviews the necessary contents for tackling the exercises proposed. These videos have been designed with Adobe Flash © (Adobe, 2009) and Apple QuickTime Pro © (Apple, 2008).

The system is made up of (Figure 6):
1. Data base, which contains the user data and collects the results of the exercises.
2. Web-based questionnaires, in which students complete exercises sets (there is a questionnaire for each level).
3. Web. This environment is written in Html language and ASP code, using the “Universal iPhone UI Kit” framework (Martin-Lafuente, 2009), which gives us an interface that is optimised for iPhone or iPod Touch type devices (see Figure 7 and Figure 8).
The questionnaires and the data base have been designed with the Web Quiz XP commercial software (Smart Lite Software, 2009). The web environment and the exercises have been optimised for a 3.5 inch screen size. AutoCad© (Autodesk, 2009) and Adobe Flash © (Adobe, 2009) were used for designing and colouring the questionnaire images.

<table>
<thead>
<tr>
<th>MODULE 1: Building with blocks</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
<th>LEVEL 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many cubes are missing?</td>
<td>10 exercises</td>
<td>Which of these isometric sketches matches the coded plan drawing?</td>
<td>Is this possible?</td>
<td>15 exercises</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A, B, C or D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MODULE 2: Identification of Sides and Views</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
<th>LEVEL 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which is the red side?</td>
<td>1, 2, 3, 4…</td>
<td>Which is the view shown?</td>
<td>Top, front or right</td>
<td></td>
</tr>
<tr>
<td>Which is the view shown?</td>
<td></td>
<td>Which piece matches the view shown?</td>
<td>1 to 6</td>
<td></td>
</tr>
<tr>
<td>Which piece matches the isometric sketch shown?</td>
<td></td>
<td>Which view matches the isometric sketch shown?</td>
<td>A, B or C</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MODULE 3: Object Discrimination</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
<th>LEVEL 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which objects match these views?</td>
<td>1 to 12</td>
<td>Which object does this match?</td>
<td>1 to 9</td>
<td></td>
</tr>
<tr>
<td>Which object do these views match?</td>
<td></td>
<td>Which object do these views match?</td>
<td>A, B, C or D</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MODULE 4: Rotations</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
<th>LEVEL 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which view do I see if I rotate the object 90º?</td>
<td>A, B or C</td>
<td>What view do I see if I rotate the object 90º?</td>
<td>A, B or C</td>
<td>How far do I have to rotate the object about the given axis to get the view shown in the second object? 90º, 180º, 270º</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MODULE 5: Cross sections</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
<th>LEVEL 4</th>
</tr>
</thead>
</table>

Table 1. Summary of course modules and levels
Hardware and Software

Although both the material and the study have been designed and implemented on iPod Touch and iPhone devices, we must highlight the fact that the course is accessible from any PC, PDA or from any mobile device in general that has at least a 3.5 inch screen size and preferably with a touch screen (connection via Wi-Fi, GPRS, 3G, etc.). At the beginning of this research activity (October of 2008), Apple’s iPod Touch was selected as the most cost effective solution to verify with requirements needed: big tactile screen (3.5 inch) and Wi-Fi access. The device’s internet browser must be HTML, CSS, ASP code compatible. The course is available on the website of the Dehaes Research Group as a moodle course and packaged as SCORM.

Pre- and post-test

Each participant has completed the Mental Rotation Test (MRT) (Albaret & Aubert, 1996) (Figure 9), before and after undergoing this experience, with the objective to be able to assess the increase in their spatial skills caused by the course. It contains 20 items divided into two sets of ten. Each item consists of a display figure and four additional
ones, two of which match the original figure after being rotated to a certain degree, the other two do not match the original display figure. The time limit is 6 minutes, divided into two periods of 3 minutes for every 10 items. It is interesting to emphasise that the results obtained by Vanderberg & Kuse (1978) clearly indicate a difference between men and women, irrespective of their ages.

Figure 9. Example of MRT test question

Computer experience and learner satisfaction survey

After completing the post-course test (MRT), participants answered some questions about their experience. Students were also asked to rate their satisfaction with the teaching content and with the use of the Ipod using a five-step Likert scale.

Procedure

This study uses two working groups: an experimental group and a control group. The experimental group is made up of 38 students who volunteered to undergo the training programme and a control group of 30 students, which undergoes no training during the week that this study lasts. The training programme has been developed in a non face to face format, for one week. Each day, the students did one of the five modules proposed (from Monday to Friday) and then they had two additional days to complete the exercises that they were unable to do on the date set for whatever reason. The week before training students do the pre-course test and fill in a user data survey. The experimental group is called to meet in a university classroom. The course programming is explained to them and they do a short practise session to familiarise themselves with the device. Table 2 shows the activities which student attend on the course for both the experimental group and the control group.

Table 2. Summary of activities

<table>
<thead>
<tr>
<th>Week</th>
<th>Day</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Monday Pre-testing and user data survey (experimental and control group)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Friday Presentation of the devices and explanation of the course programme (experimental group)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Monday Module 1: Building with blocks</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Tuesday Module 2: Identification of Sides and Views</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Wednesday Module 3: Object Discrimination</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Monday Module 4: Rotations</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Friday Module 5: Cross Sections</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Saturday Complete levels/modules not previously completed if necessary</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Sunday Complete levels/modules not previously completed if necessary</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Tuesday Post-testing (experimental and control group)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Administration of learner satisfaction survey (experimental group)</td>
</tr>
</tbody>
</table>

Data Analysis and Results

Prior Spatial Training

The pilot study was carried out with 68 first year engineering students, in October, at the beginning of the first term of the 2008-2009 academic year. This way, we tried to avoid interference from other subjects on the course on the results of the spatial visualisation test measurements. The participants’ levels of spatial visualization (SV) were measured by using the Mental Rotation Test prior to spatial training (pre-course test). The scores shown in the tables
are for the number of correct answers, from a maximum possible of 40 points in this test. The mean score of the pre-course test for females (n=23) and males (n=45) were 14.03 and 20.02 respectively (see Table 4).

Differences attributed to “group” or “gender” factors in spatial visualisation were investigated prior to the treatments of spatial training. Gender is one of the variables considered in the majority of studies about spatial abilities. A two-way analysis of variance (2-way ANOVA) with interactions was performed to detect if there were any gender or group differences among the participants in the study. For the Mental Rotation pre-course test, there were no significant differences between these groups (experimental vs control) prior to spatial training, $F(1,64) = 1.254, p = 0.267$. Concerning gender, the differences in the average scores between men and women were statistically significant, $F(1,64) = 8.907, p = 0.004$. However, the interaction between gender and group is not shown to be significant, $F(1,64) = 1.019, p = 0.317$. Table 3 summarises the results of the variance analysis for the pre-MRT.

Summing up, we can say the analysis shows no significant difference between the two groups from this study, however there are mean differences between gender, which agrees with the studies in the scientific literature on this subject that point to a difference in spatial visualisation levels in men and women (Vanderberg & Kuse, 1978).

**Table 3. ANOVA results for MRT pre-course test by training condition and gender and their interactions**

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$ Value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>75.582</td>
<td>1</td>
<td>75.582</td>
<td>1.254</td>
<td>.267</td>
</tr>
<tr>
<td>Gender</td>
<td>536.788</td>
<td>1</td>
<td>536.788</td>
<td>8.907</td>
<td>.004*</td>
</tr>
<tr>
<td>Group*Gender</td>
<td>61.420</td>
<td>1</td>
<td>61.420</td>
<td>1.019</td>
<td>.317</td>
</tr>
<tr>
<td>Error</td>
<td>3857.019</td>
<td>64</td>
<td>60.266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4489.941</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p <.01

### After Spatial Training

For the statistical analysis, we used a two-way analysis of covariance (2-way ANCOVA), taking as the null hypothesis ($H_0$) the fact that mean values for spatial visualisation did not vary after the end of the course. The post-MRT has been used as the dependent variable, the pre-MRT as the co-variable and the group and gender as independent variables. Moreover, different interactions between the factors (group, gender) and between the factors and the co-variables (pre-MRT) have been calculated, leaving the pre-MRT*Group interaction in the model. Table 4 shows the mean scores and the standard deviations by group and sex, obtained by the participants in the pre-course test (pre-MRT), in the post-course test (post-MRT) and the mean gain found (gain-MRT).

**Table 4. Mean and Standard Deviation measures for pre- and post-course test and gain scores**

<table>
<thead>
<tr>
<th>Group</th>
<th>Gender</th>
<th>N</th>
<th>pre-MRT (std. dev)</th>
<th>post-MRT (std. dev)</th>
<th>gain-MRT (std. dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Males</td>
<td>19</td>
<td>19.89 (10.03)</td>
<td>22.37 (9.96)</td>
<td>2.47 (2.84)</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>11</td>
<td>11.91 (5.24)</td>
<td>12.91 (5.15)</td>
<td>1.00 (2.05)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>30</td>
<td>16.97 (9.34)</td>
<td>18.90 (9.60)</td>
<td>1.93 (2.64)</td>
</tr>
<tr>
<td>Experimental</td>
<td>Males</td>
<td>26</td>
<td>20.12 (7.23)</td>
<td>28.54 (6.98)</td>
<td>8.42 (7.80)</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>12</td>
<td>16.17 (6.51)</td>
<td>23.33 (8.96)</td>
<td>7.17 (5.52)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>38</td>
<td>18.87 (7.17)</td>
<td>26.89 (7.93)</td>
<td>8.03 (7.11)</td>
</tr>
<tr>
<td>Total</td>
<td>Males</td>
<td>45</td>
<td>20.02 (8.42)</td>
<td>25.93 (8.82)</td>
<td>5.91 (6.83)</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>23</td>
<td>14.13 (6.20)</td>
<td>18.35 (8.97)</td>
<td>4.22 (5.21)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>68</td>
<td>18.03 (8.19)</td>
<td>23.37 (9.52)</td>
<td>5.34 (6.34)</td>
</tr>
</tbody>
</table>

* Values are expressed as a mean (standard deviation).

The ANCOVA analysis is summarised in Table 5. Figure 10 is a graphic presentation of the means obtained by the groups. The results are shown separately for males (M) and females (F). The analysis found a positive relationship between the pre-MRT and the post-MRT, $F(1,63) = 81.130, p < 0.001$, and there were no significant differences between genders $F(1,63) = 3.088, p = 0.084$. Significant differences were found between the experimental group and
the control group, $F(1,63) = 16.745, p < 0.001$, with these varying in accordance with the pre-MRT scores, as can be seen in the significant interaction Group*pre-MRT, $F(1,63) = 5.078, p = 0.028$.

The multiple regression model obtained in this analysis is shown in Equation 1. Gender is not significant at an alpha level of 0.05, although p-value is 0.084 (less than 0.1 as seen on Table 5), and it must be emphasized that there is a mean difference of 2.502 points advantage score for males. For this reason, it has been considered including gender in the regression model. Figure 11 shows how individuals with high initial scores have little margin for improving them meanwhile students with low initial scores have greater improvement chances (see Figure 11) according to findings obtained in other research works.

**Table 5. ANCOVA results for post-MRT by group and gender and their interactions**

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>454.957</td>
<td>1</td>
<td>454.957</td>
<td>16.745</td>
<td>.000**</td>
</tr>
<tr>
<td>Gender</td>
<td>83.911</td>
<td>1</td>
<td>83.911</td>
<td>3.088</td>
<td>.084</td>
</tr>
<tr>
<td>Pre-MRT</td>
<td>2204.208</td>
<td>1</td>
<td>2204.208</td>
<td>81.130</td>
<td>.000**</td>
</tr>
<tr>
<td>Group*pre-MRT</td>
<td>137.962</td>
<td>1</td>
<td>137.962</td>
<td>5.078</td>
<td>.028*</td>
</tr>
<tr>
<td>Error</td>
<td>1711.647</td>
<td>63</td>
<td>27.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6069.809</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

**Figure 10.** Mean scores in pre-course and post-course tests for the Experimental Group and the Control Group (N=68)

\[
\text{post-MRT} = 14.297 + 2.502 \times (\text{gender=male}) - 12.826 \times (\text{group=control}) + 0.577 \times \text{pre-MRT} + 0.357 \times (\text{group=control}) \times \text{pre-MRT}
\]

*Equation 1. Multiple regression model*

After the detailed statistical analysis shown below, concerning spatial training, this web-based course, assessed on iPod Touch devices, we may conclude that:

- For the initial values of spatial visualisation, measured with the MRT test (pre-MRT), women have a lower mean score than men (see Figure 10). This matches with the initial studies by creators of this test and with other papers from the scientific literature (Rafi, Samsudin & Said, 2008; Martin-Dorta, Saorin & Contero, 2008).
- Moreover, results show that, with specific training, men and women achieve a similar mean gain in the MRT test. Despite the initial differences between men and women, both genders manage increasing their score in the Mental Rotation Test by around 8 points (see Figure 10).
- The main effect of treatment group (experimental group vs control group) was significant, $F(1,63) = 16.745, p < 0.001$, showing better performances for participants receiving training compared to the control group. The mean gain score was around 8 points (see Figure 10).
Some researchers establish 60% of correct answers (around 24 points out of 40) as the minimum level of spatial ability for freshmen engineering students (Sorby, 2007). The students in this study achieved an average score around 26 points (see Table 4 and Figure 10).

That mean gain obtained by the experimental group, in comparison with the control group, depends on the value of the pre-MRT. For example, if initial score in the test (pre-MRT) of an individual is 10 points, the mean gain made is 9.256 points greater in the group that undergoes training in comparison with the individuals that do not receive spatial training. If score obtained by an individual in the pre-MRT were 30 points, the mean gain would reduce to 2.116 points. These results are according to findings obtained in other research works: individuals with high initial scores have little margin for improving them. The students with low initial scores have a greater chance to improve (see Figure 11 and Equation 1).

Figure 11. Mean scores in pre-course and post-course tests for the Experimental Group and the Control Group (N=68)

Students’ satisfaction survey

The students’ satisfaction survey was structured in two parts: a first part focussing on satisfaction with the device and a second part that assessed general satisfaction with the course. Table 6 shows the results of the survey in which users evaluate their experience with the iPod, scoring it on a Likert scale, which ranges from very bad to very good (1=VERY BAD, 2=BAD, 3=UNDECIDED, 4=GOOD, 5=VERY GOOD). Two students weren’t able to finish the survey.

<table>
<thead>
<tr>
<th>Item n°</th>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How do you rate your experience with the touch screen?</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>19</td>
<td>15</td>
<td>4.36</td>
<td>0.59</td>
</tr>
<tr>
<td>2</td>
<td>How do you rate your experience with the touch keyboard?</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>19</td>
<td>8</td>
<td>3.92</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td>How do you rate the device’s Wi-Fi connection?</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>16</td>
<td>3</td>
<td>3.42</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>How do you rate the internet navigation experience?</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>22</td>
<td>4</td>
<td>3.81</td>
<td>0.67</td>
</tr>
<tr>
<td>5</td>
<td>Overall, How do you rate your general satisfaction with the device?</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>10</td>
<td>4.08</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Overall satisfaction with the device used was rated by 91.43% of the users as GOOD or VERY GOOD. The touch screen and internet navigation were the best rated ($m = 4.36$ and $m =3.92$ respectively), while Wi-Fi network was the worst rated ($m = 3.42$). We may highlight that several students pointed out some problems with their domestic networks.
Students’ attitudes towards learning with web-based course material are presented in Table 7. A five-step Likert scale was used in this study to assess user satisfaction with the course (1=STRONGLY DISAGREE, 2=DISAGREE, 3=UNDECIDED, 4=AGREE, 5=COMPLETELY AGREE).

Nearly 90% of students scored the presentation and the structure of the teaching material used on the course between 4 and 5 points on a Likert scale of five points (m = 4.28 and m = 4.19 respectively). Moreover, they were asked to indicate the exercises that, in their opinion, presented the greatest difficulty. They consider that Level 3 of Module 1 is one of the most complicated ones, followed by modules 4 and 5. But around 90% of them stated that they were capable of doing all the tasks posed (m = 4.22).

Around 94% agree that the course offers useful contents (m = 4.11) and 75% of them would like having this kind of teaching material for support in the subjects of their degree course (m = 3.97). 80.55% agree or completely agree that it is important to have access to contents of this kind where and whenever they like (m = 4.19).

Regarding format preferences while doing this course, we would highlight the fact that only 26.47% would agree or would completely agree to do this course using the traditional format of pencil and paper (m = 2.56). Around 90% of them would prefer to do it on a computer and 80% of them prefer to do it on handheld devices (mobile phones, consoles, etc.) (m = 4.47 and m = 3.97 respectively).

We may emphasize that only 5.56% (m = 2.42) of them would agree to do this course with the physical presence of the teacher.

### Table 7. Students’ answers to attitude questionnaire regarding the web course experience (N=36)

<table>
<thead>
<tr>
<th>Item n°</th>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The didactic content of the course was well and carefully presented</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>11</td>
<td>4.28</td>
<td>0.51</td>
</tr>
<tr>
<td>2</td>
<td>The structuring of the course in modules and levels is right</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>21</td>
<td>11</td>
<td>4.19</td>
<td>0.62</td>
</tr>
<tr>
<td>3</td>
<td>I felt capable of solving the exercises</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>11</td>
<td>4.22</td>
<td>0.59</td>
</tr>
<tr>
<td>4</td>
<td>I consider the videos are a good tool for assimilating the contents</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>11</td>
<td>3.89</td>
<td>0.91</td>
</tr>
<tr>
<td>5</td>
<td>The course offers useful contents</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>28</td>
<td>6</td>
<td>4.11</td>
<td>0.46</td>
</tr>
<tr>
<td>6</td>
<td>I would like to have this kind of material as a teaching support in face to face teaching, both in this subject and in others</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>15</td>
<td>12</td>
<td>3.97</td>
<td>0.82</td>
</tr>
<tr>
<td>7</td>
<td>It is important to have access to the contents of this course when and wherever I want</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>13</td>
<td>16</td>
<td>4.19</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>If you had to choose between different formats for doing this course, which one would you prefer?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pencil and paper</td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>2.56</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>PC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>17</td>
<td>4.47</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Handheld devices</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>18</td>
<td>10</td>
<td>3.97</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>I think that the use of the iPod, with its touch screen technology, for doing on-line courses is a good learning system.</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>16</td>
<td>17</td>
<td>4.39</td>
<td>0.64</td>
</tr>
<tr>
<td>10</td>
<td>I would have liked to have done this course with the teacher physically present</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>13</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>2.42</td>
<td>0.81</td>
</tr>
<tr>
<td>11</td>
<td>Would you recommend this course to your colleagues?</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>27</td>
<td>8</td>
<td>4.19</td>
<td>0.47</td>
</tr>
<tr>
<td>12</td>
<td>Overall, I am satisfied with this course</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>9</td>
<td>4.25</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The students also gave the course an overall rating where 75% of them considered that they were satisfied with the course (4 points) and the remaining 25% were totally satisfied (5 points) (m = 4.25). Moreover, almost all students would recommend the course to their colleagues (m = 4.19).

### Limitations and Conclusions

Some limitations of the present research have to be taken into account. In the present study, we didn’t have in mind some variables that may have an effect on results, such as previous experience with touch screen devices, motivation,
age and previous experience in engineer graphic course or with CAD software. In further investigations that variables may be used.

Data obtained from user satisfaction survey allows us stating that this interactive web-based course done on handheld touch screen devices motivated the students that took part in it, generating a high degree of satisfaction, and the participants stated that they would recommend it to their colleagues/friends. This is a very important factor making students willing to take part in it, such as our short remedial course.

The didactic content designed for this course was user friendly and they rated browsing with device’s touch screen very positively (>97%). It seems worth emphasizing the fact that they stated they prefer doing a course of this kind without the physical presence of the teacher, which could indicate that this online learning format and its non-formal use helps young users choosing it as it suits regular leisure habits of today’s young people (computer games, video games consoles).

They also rate positively access to contents from any place at any time (>80%) and they are not quite receptive to any courses given or done in the traditional way of pencil and paper (<27%). Furthermore, the course has shown that it is a good choice for developing spatial visualization. The statistical analysis shows a significant increase of around 8 points in the Mental Rotation Test.

It is important emphasizing that this course offers a new context for developing spatial abilities: in first place it offers the use of a new hardware device, the handheld touch screen devices, while it also offers mobility for the teaching-learning process and a distance learning course which can be used on the subway, bus, tram or any place where a wireless internet access is available (Wi-Fi, 3G, GPRS, etc.). There are many studies and developments done for spatial training, but it is the first that develops application with mobile devices.

This work has been led by interest in developing material that includes development of spatial visualisation related skills, promulgated by the EHEA and included in the new Engineering degree courses offered by the Spanish Ministry of Education. We believe it is necessary keeping development of material of this kind that provides students with mobility opening up new teaching and learning possibilities according to actual students habits and lifestyle.

Acknowledgements

This work is supported by the “Program for Studies and Analysis” (Project “Evaluation and development of competencies associated to the spatial ability in the new engineering undergraduate courses” ref. EA2009-0025) funded by the Spanish Ministry of Education (ME).

References


Using Reflective Peer Assessment to Promote Students’ Conceptual Understanding through Asynchronous Discussions

Huann-shyang Lin¹, Zuway-R Hong², Hsin-Hui Wang² and Sung-Tao Lee³

¹Center for General Education, National Sun Yat-sen University, Kaohsiung, Taiwan // ²Institute of Education, National Sun Yat-sen University, Kaohsiung, Taiwan // ³Department of Applied Science, Naval Academy, Kaohsiung, Taiwan // huannlin@faculty.nsysu.edu.tw // a3803429@ms49.hinet.net // hsinhui5885@gmail.com // sungtao@mail.cna.edu.tw

ABSTRACT
This study explores the impact of using assessment items with competing theories to encourage students to practice evaluative reflection and collaborative argumentation in asynchronous discussions. Thirty undergraduate students from various departments worked in small groups and took turns collaboratively discussing the given item’s answer, reaching a consensus, and posting their consensual answer on the web. The remaining participants served as evaluators to reflect on the answer and provide comments. It was found that the students made significant progress in argumentation ability and conceptual understanding of related scientific content knowledge. In the beginning of the study, the group of students majoring in the sciences outperformed counterparts with non-science majors on the level of understanding of the assessment item’s scientific concepts. At the end of the semester, the differences diminished between the two groups both on conceptual understanding and in argumentation ability.

Keywords
Argumentation, Asynchronous, Conceptual understanding, Reflection

Introduction
Existing literature has indicated that online asynchronous discussion has the potential to generate the critical dimensions of learning found in traditional classrooms (Andresen, 2009); furthermore, it has the ability to enhance higher cognitive levels of knowledge construction (Schellens & Valcke, 2005). However, simply putting students in an asynchronous discussion environment does not necessarily bring about collaborative interactions and effective outcomes, because some students may be reluctant to disagree with others (Andriessen, 2006) or will vary in their level of involvement (Veerman, 2003). In order to promote the effectiveness of asynchronous discussions, pioneers in collaborative learning examined a variety of strategies. For example, assigning roles to students at the beginning of discussions resulted in a significant positive impact on students’ level of knowledge construction (De Wever, Van Keer, & Valcke, 2009); Elaborating on the meaning of discussion questions promoted balanced argumentation for all participants, especially for those with less knowledge of the question (Golanics & Nussbaum, 2008); and engaging students in reflective interactions such as explaining, justifying, and evaluating problem solutions has shown a productive learning outcome for physics modeling tasks (Baker & Lund, 1997). After considering the strategies recommended by existing literature, we designed an assessment instrument with competing theories for students to practice evaluative reflection through asynchronous discussions and examined its impact on key competencies of educational outcomes: students’ conceptual understanding and argumentation ability (Driver, Newton, & Osborne, 2000).

Argumentation in asynchronous discussion environment
Social constructivism theory furthers the idea that learning effect can be promoted through active interactions and communication among participants. For social constructivism, knowledge is created and legitimized by means of social interactions between and among individuals in a variety of community, societal, and cultural settings (Driver, Leach, Millar, & Scott, 1996; Staver, 1998). When this theory is appropriately applied in the context of education, students are encouraged to interact with others to construct individual understanding and knowledge. It also provides opportunities for students to reflect on other classmates’ comments, suggestions, presentations, and ways of learning.

In the process of online argumentation, students are encouraged to actively write, discuss, and debate online using text-based communication tools. Students make progress in argumentation by providing evidence-based conclusions, describing why they agree or disagree with the presented statements, and try to persuade others. In order to provide
quality arguments, students must explain their own positions, evaluate current arguments, summarize peer comments, and integrate related information or knowledge. All of these activities seem to be helpful for students’ clarification of conceptual understanding and improvement of argumentation. More importantly, the time delays in text-based asynchronous discussions provide opportunities for students, especially for those who need more time, to reflect and scrutinize online information (Veerman, Andriessen, & Kanselaar, 2000). In collaborative learning, meaning is produced by examining the relationship between utterances through social interactions; meaning is examined and reconstructed as a direct result of conflict or argumentation in a social context (Jeong & Joung, 2007). For critical argumentation, students can use “counter-arguments” to “challenge” other students’ statement when they disagree with the statement (Veerman et al., 2000). The learning of argumentation is consistent with the development of scientific knowledge in most science communities. Vigorous discussions, debates, and peer-review procedures are exchanged among scientists. Scientists then provide counter-arguments or rebuttals to challenge the data, evidence, or assertions of other scientists with different theories. For example, in the eighteenth century, the phlogiston theory was almost universally accepted at the time and was the basis of the chemistry taught to college students then. The theory hypothesized that during combustion the substance of phlogiston was released and combined with air (Conant, 1957; Harre, 1981). However, French scientist Lavoisier provided empirical evidence of burning mercury and phosphorous to challenge the phlogiston theory. He found that the result of burning mercury or phosphorous did not decrease their weights as the theory predicted, instead, the final weights increased. This rebuttal and more empirical data from his follow-up experiments (e.g., an experiment collecting the gas formed after heating the red oxide of mercury) play key roles for the demise of the phlogiston theory. By reviewing the development of scientific knowledge and the history of science, we can find that argumentation plays a central role in the resolution of scientific controversies (Fuller, 1997).

Similar to the development of scientific knowledge, argumentation deserves a place in the pedagogy of science. In arguing the use of argumentation theory in education, Driver et al. (2000) concluded that helping students to construct coherent arguments and evaluate others are important skills, especially pertaining to topics reported in the media. This is even more so in our contemporary and democratic society, since there are many public policies relating to science and the public has a legitimate voice (e.g., use of bio-ethanol as fuel, restriction of genetically modified foods, and control of air quality) (Newton, Driver, & Osborne, 1999). Through the practices of posing and evaluating arguments, students become active participants in the learning community rather than just passive knowledge receivers. Another potential benefit of collaborative argumentation in stimulating students’ conceptual understanding and belief revision has been examined (Ravenscroft, 2000). In the study, the learner adopts the role of an explainer while the computer system plays a facilitating role; participants collaborate to develop a shared explanatory model. Ravenscroft found that students revised their beliefs and improved their explanatory models in a delayed post-test.

Despite the emphasis of argumentation ability in science teaching, it is rarely adopted in typical classroom teaching. Major reasons ranged from teachers’ perception of the difficulty and challenge in managing group discussions to the time pressure imposed by the need to cover the national curriculum (Driver et al., 2000; Newton et al., 1999). The literature review reveals that there are limitations and constraints for teachers to implement collaborative argumentation or group discussion in typical classroom teaching practices. On the other hand, benefits of online asynchronous discussion may make students’ interactions and group discussions more effective than traditional face to face discussions. For example, it allows slow-paced students more time to construct arguments and contribute to the discussion (Veerman et al., 2000); the transcript of the discussion is always available for participants’ reference (Weinberger & Fischer, 2006); and discussions generally will not be interrupted by a particularly aggressive participant (Andriessen, 2006). Furthermore, in typical classroom face to face settings, most of the interactions are generally dominated by outspoken students. The learning opportunity of practicing communication or argumentation for students who are either shy or weaker speakers is unintentionally limited or deprived (Nussbaum & Jacobson, 2004). Therefore, the discourse of the science classrooms needs to be more deliberative or dialogic (Simon, Erduran, & Osborne, 2006). This situation of inequity in learning deserves attention from teachers and educators. Although there is reason to hypothesize that an asynchronous discussion could be one of the effective alternatives to conduct argumentation activities, more empirical studies focusing on inspiring the practice of reflection and promoting the highly expected educational outcome of aforementioned argumentation and conceptual understanding are needed. These studies should also aim to enable us to better understand what role computers can play in supporting the classroom teaching that has mostly failed to promote these higher level cognitive abilities (Webb, Jones, Barker, & van Schaik, 2004).
The importance of reflective ability

Reflective ability has long been regarded as one of the major goals for students’ learning outcome. Early in 1933, Dewey proposed reflection as a process of problem solving. Recently, the Organisation for Economic Co-Operation and Development (OECD, 2006) officially identified reflective ability as an important component of scientific, reading, and mathematical literacy. Based on the OECD’s definition, the competency of reflection is deemed to be essential for an improved future. Knowing the importance of reflective ability, researchers and educators have asserted that it can be taught and trained through proper design of activities (Boud, Keogh, & Walker, 1985; Lee & Hutchison, 1998). An investigation of teaching effectiveness by designing a variety of online activities has been conducted. For example, Chen et al. (2009) used “reflection prompts” to engage online learners in a reflective learning environment. They found that the reflection level of those students who had been provided with “high level prompts” was significantly higher than the level of those who were not shown these prompts. Review of the literature above reveals that providing reflective learning opportunities for students could be fruitful in promoting expected outcomes. Further examination of current practices of reflective activities found that most studies mainly focused on the assessment of reflective ability level (Chen et al., 2009; Gulberg & Pilkington, 2007; Yang, 2009). A limited number of studies have investigated how reflective activities influenced learners’ key competencies in educational outcomes (e.g., argumentation or science concept understanding). To this end, we designed a semester-long online reflective peer assessment and investigated its impact on students’ argumentation and conceptual understanding.

Typical assessment vs. Reflective peer assessment

The term of reflective peer assessment means in a collaborative learning environment, students critically assess each other’s feedback posted online, vigorously discuss various perspectives, and continuously reflect and elaborate on their own assertions (Veerman et al., 2000). For typical classroom teaching practices, although inquiry skills or group discussions can be used to promote student-teacher or student-student interactions to assess student conceptual understanding (Van Zee, Iwasyk, Kurose, Simpson, & Wild, 2001), time constraints or lack of professional ability may prevent instructors from employing these teaching strategies (Newton et al., 1999; Roth, 1996). In this study, we propose a reflective peer assessment in an asynchronous discussion learning environment to promote collaborative learning and critical argumentation. The details of reflective peer assessment will be explained with examples in the methodology section.

Although considerable research has been devoted to computer supported collaborative learning (CSCL) through the perspectives of investigating collaborative behaviors (Hsu, Chou, Hwang, & Chou, 2008), identifying strategies and factors related to better collaboration (De Smet, Van Keer, & Valcke, 2008; Onrubia & Engel, 2009), and analyzing the effect and the role of peer feedback (Lai & Lan, 2006; Tseng & Tsai, 2007), rather less attention has been paid to the systematic integration of key competencies of educational outcomes in the CSCL environment: reflection, argumentation, and conceptual understanding.

Purpose of the study

As mentioned earlier, interactions of student learning among the three variables of reflection, argumentation, and conceptual understanding have rarely been investigated simultaneously. A literature review reveals that asynchronous discussions provide opportunities for students to work collaboratively and have great potential in promoting learning outcomes, especially when students were encouraged to reflect on another team’s consensual answer and peer arguments and the reasons why they agree or disagree with the consensual answer. The purpose of this study is to explore the impact of using reflective peer assessment in asynchronous discussions on the development of students’ argumentation ability and conceptual understanding. Specifically, the research questions are as follows:

1. How students progress on the development of their argumentation ability and conceptual understanding through the six rounds of reflective peer assessment in asynchronous discussions?
2. Do both science major and non-science major students benefit from reflective peer assessment over the duration of the study?
Method

This study was conducted in the context of an undergraduate course—history of science. The reflective peer assessment instrument, participants, procedure, and data analysis are described as follows:

Reflective peer assessment instrument

The reflective peer assessment instrument contained six open-ended question items. We developed four items related to gas laws and buoyancy. The four items were validated by three science educators and two physical science teachers who were asked to judge each item by the following criteria:
1. The item examines the conceptual understanding and/or application of (Boyle’s law, Charles’s law, atmospheric pressure, or buoyancy).
2. The item is compatible to the topics that students have previously studied in high school physical science.
3. The item is clearly phrased.

Two more items were derived from the released item bank of Programme for International Student Assessment (PISA) 2006 (OECD, 2007) requiring students to provide evidence-based conclusions. The six-item instrument was pilot tested for one year prior to the study and found to be reasonably reliable (Cronbach $\alpha=0.72$). In order to check for students’ conceptual development, two items with a similar difficulty level related to gas laws were selected for comparison—one was randomly assigned and assessed at the beginning of the semester while the other item was measured at the end of the course.

In order to promote students’ argumentation discussions, we followed the suggestion of Osborne, Erduran, and Simon (2004) to integrate competing theories into the assessment items. For each item, students are asked to make an assertion or take a position and are encouraged to provide persuasive arguments with appropriate theory, principle, reasonable data or evidence, and supportive warrant or backing. Whenever there are disagreements or different positions of the posted statements, students are encouraged to use rebuttals to challenge the existing statements. One sample item can be seen in the Appendix. For this item, the scientific claim for the first question should be “the mercury moves to the left-hand side”. For the second question of the item, reasonable arguments should be similar to the following explanation containing sound conceptual understanding, basic comprehension of data (relating variables of volume, temperature, and pressure), and the backing (with theories or principles) of an argument: “The air inside the flask enclosed by the mercury is a closed system. In the beginning status of 25° C, its pressure is equal to the surrounding atmospheric pressure. When the set of the flask is moved to the outdoor 5° C environment, the surrounding pressure (i.e., atmospheric pressure) stays the same. Therefore, the volume inside the closed system decreases as the temperature decreases from 25° C to 5° C. This is based on Charles’s Law which states that when pressure is kept constant, a certain amount of gas volume ($V$) is proportional to the temperature ($T$) (i.e., $V_1/T_1=V_2/T_2$). For this question, high quality or advanced arguments even provide rebuttals or counter-arguments to challenge the statements with different claims or predictions.

Participants

The 30 participants (21 males and 9 females) of the study were undergraduate students from colleges of art and humanities, sciences, engineering, management, and social science. They were enrolled in the course entitled History of Science and invited to participate in the online asynchronous discussions. Their ages ranged from 20 to 24 years old. It should be noted that the participants were assumed to have learned all of the relevant content (i.e., Boyle’s law, Charles’s law, atmospheric pressure, greenhouse effect, and buoyancy) related to the six test items in their high school physical science course and no further scientific concepts were instructed during the study. For the purpose of promoting collaborative interactions in solving a given item, the students were randomly divided into six groups. One of the six groups was asked to collaboratively discuss the questions of the given item, reach a consensus, and post their consensual answer on the web. A screenshot of the discussion system is shown in Figure 1. The rest of the students were asked to serve as evaluators for the statement posted by the group.
The role of the evaluator is to reflect on the posted statement, analyze the level of statement using the model developed by Osborne and his colleagues (2004), check the correctness of the statement, assign an appropriate score from 1 to 5 where 1 is the lowest quality while 5 stands for the highest quality of argumentation level for the posted statement, and explain the details of evaluation, reflection, or rebuttals.

Procedure

In the beginning weeks of the course, the instructor explained the difference between scientific arguments and personal opinions, described the elements of an argument with examples, introduced the role of online discussions, and assigned the sequence and date for each group of students to work either as poster (of an item’s answer) or as evaluators. We use Toulmin’s argumentation model (Toulmin, 1958) to introduce the essential elements—data, warrant, backing, and rebuttal and follow the recommendations of Osborne’s research team on the evaluation of arguments (Osborne, Erduran, & Simon, 2004b). In the fourth week, the first group was asked to respond to item 1 and post their consensual answer on the web within one week. During the fifth week, the rest of the students were asked to evaluate the posted answer individually. The evaluators were free to revise their comments according other evaluators’ feedback—allowing for a dynamic process of reflection. In the same time, the group members who posted the answer continued to reflect on the evaluators’ feedbacks and arguments either support or against their answer. At the end of the fifth week, they were asked to present their final answer and justification of their position in a class meeting. For each of the rest of the 5 argumentation items, the procedure was similar to item 1; that is, in the first week the assigned group was responsible for answering the item, while the rest of the students served as reflective evaluators responsible for posting personal comments within the following week. In total, the asynchronous discussions of reflective peer assessment lasted for 12 weeks for the six items.

In order to promote students’ involvement in argumentation, all six items asked students justify their position or explain their reason. In the beginning of the semester, examples of high level arguments were explained to the students for the purpose of scaffolding their ability in argumentation. Through analyzing their arguments in six
rounds of reflective peer assessment, we hope to gain insights that inform subsequent initiatives aimed at a wider application of asynchronous discussion in the development of high level cognitive abilities.

**Data analyses**

Students’ conceptual understandings were examined and analyzed verbatim based on their reasoning arguments, explanations, and comments. The number of alternative conceptions, the level of conceptual understanding, and the quality level of students’ arguments were used as quantitative indicators.

The scoring scheme for students’ level of conceptual understanding is based on our previous studies (Lin, Cheng, & Lawrenz, 2000; Lin, Hung, & Hung, 2002). The scheme gives 3 points to the answers with correct statements and use of target scientific concepts (e.g., for the sample item, appropriately use the key concepts and identify that in a closed system with constant pressure the gas volume inside the system would change proportionally with the surrounding temperature); 2 points for those answers with sound explanations but minor mistakes (e.g., unable to identify any one of the above key concepts. For instance, fail to explain that the air inside the flask enclosed by the mercury is a closed system); 1 point for the statements showing partial misconceptions but indicating some degree of relevance toward the target concept (e.g., refer to the test item is related to Boyle’s law or Charles’ law but fail to explain how the laws can be used in the item); 0 points for no explanation or explanations with irrelevant statements or misconceptions.

For the scoring scheme on the level of students’ arguments, we follow the method developed by Osborne, Erduran, and Simon (2004), in which, level 1 arguments are arguments with a simple claim without containing any other elements (e.g., I thought that they explained it very well. Without reviewing the rebuttals posted by others, it is difficult for me to identify the weakness of the answer.); level 2 arguments consist of claims with either data, warrants, or backing but do not contain any rebuttals (e.g., this group appropriately used Boyle’s law to explain the mercury movement. However, the lack of real life examples makes it not persuasive); level 3 arguments consist of a series of claims or counter-claims with either data, warrants or backings with the occasional weak rebuttal (e.g., I think there is little difference between the indoor and outdoor air pressure. The major influence for the mercury’s movement should be from the volume change of the enclosed gas.) ; level 4 arguments consist of a claim with a clearly identifiable rebuttal (e.g., Moving the whole set from 25°C indoors to 5°C outdoors at a constant pressure of 1 atm, the mercury would not move to the right, because Charles’ law tells us that at constant pressure, gas volume would shrink instead of expanding); level 5 arguments contain claims supported by data and warrants with more than one rebuttal (e.g., using theoretical backgrounds or evidences to refute the prediction of mercury moving to right).

Two science educators with the domain knowledge of chemistry and physics evaluated the participants’ argumentation content and statement based on the above scoring scheme. Since item 1 was used as a practice item for students, answers and comments of this item served as examples for the two evaluators to discuss the detail of the scoring procedure. After the discussion, the two evaluators scored items independently. At the end of the scoring procedure, the two evaluators discussed the statements which had discrepancies between their scoring until a consensus was reached. The mean of the two evaluators’ scores was used as the final score for each individual student’s score of the target item.

**Result**

The first research question of the study was intended to investigate student argumentation ability and conceptual understanding in an asynchronous discussion environment. Meanwhile, the second research question attempted to check the “equity” of the educational opportunity and learning environment provided by the study’s treatment.

**Students’ performance of argumentation and conceptual understanding in an asynchronous discussion environment**

Table 1 presents the means and standard deviations of the participants’ performance on the five assessment items. Since item 1 was used as practice item for students to be familiar with the procedure of asynchronous discussions, it
was not evaluated for data analysis. It can be seen from Table 1 that students made gradual progress on their argumentation ability from a mean score of 2.58 for item 2 to a mean score of 3.86 for item 6. It should also be noted that the standard deviations decreased from 1.04 to 0.33, which suggests that in the beginning, students’ argumentation ability was much more diversified than it was at the end of the study. In other words, the participants’ argumentation ability had gradually become more homogeneous. The pattern of students’ progress on their conceptual understanding is similar to the progress pattern of argumentation. The mean score improved from 1.98 for item 2 to 2.83 for item 6, while the standard deviation decreased from 0.67 to 0.34.

Table 1. Mean and sd of student performance

<table>
<thead>
<tr>
<th>Item</th>
<th>#</th>
<th>mean of argument</th>
<th>sd</th>
<th>mean of conception</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.58</td>
<td>1.04</td>
<td>1.98</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.87</td>
<td>0.97</td>
<td>2.40</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.69</td>
<td>0.47</td>
<td>2.91</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.36</td>
<td>0.49</td>
<td>2.81</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.86</td>
<td>0.33</td>
<td>2.83</td>
<td>0.34</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Comparisons between science and non-science major students

<table>
<thead>
<tr>
<th>Variable</th>
<th>group</th>
<th>mean(sd)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument 2</td>
<td>science major</td>
<td>2.93(0.86)</td>
<td>1.92</td>
</tr>
<tr>
<td>Argument 2</td>
<td>non-science major</td>
<td>2.23(1.12)</td>
<td></td>
</tr>
<tr>
<td>Argument 3</td>
<td>science major</td>
<td>3.06(1.00)</td>
<td>1.19</td>
</tr>
<tr>
<td>Argument 3</td>
<td>non-science major</td>
<td>2.64(0.93)</td>
<td></td>
</tr>
<tr>
<td>Argument 4</td>
<td>science major</td>
<td>3.80(0.41)</td>
<td>1.34</td>
</tr>
<tr>
<td>Argument 4</td>
<td>non-science major</td>
<td>3.55(0.52)</td>
<td></td>
</tr>
<tr>
<td>Argument 5</td>
<td>science major</td>
<td>3.46(0.52)</td>
<td>0.86</td>
</tr>
<tr>
<td>Argument 5</td>
<td>non-science major</td>
<td>3.27(0.47)</td>
<td></td>
</tr>
<tr>
<td>Argument 6</td>
<td>science major</td>
<td>3.85(0.34)</td>
<td>-0.15</td>
</tr>
<tr>
<td>Argument 6</td>
<td>non-science major</td>
<td>3.88(0.35)</td>
<td></td>
</tr>
<tr>
<td>concept 2</td>
<td>science major</td>
<td>2.27(0.62)</td>
<td>2.57*</td>
</tr>
<tr>
<td>concept 2</td>
<td>non-science major</td>
<td>1.68(0.61)</td>
<td></td>
</tr>
<tr>
<td>concept 3</td>
<td>science major</td>
<td>2.60(0.51)</td>
<td>3.04**</td>
</tr>
<tr>
<td>concept 3</td>
<td>non-science major</td>
<td>2.10(0.32)</td>
<td></td>
</tr>
<tr>
<td>concept 4</td>
<td>science major</td>
<td>3.00(0.00)</td>
<td>1.51</td>
</tr>
<tr>
<td>concept 4</td>
<td>non-science major</td>
<td>2.78(0.44)</td>
<td></td>
</tr>
<tr>
<td>concept 5</td>
<td>science major</td>
<td>2.88(0.35)</td>
<td>0.61</td>
</tr>
<tr>
<td>concept 5</td>
<td>non-science major</td>
<td>2.75(0.46)</td>
<td></td>
</tr>
<tr>
<td>concept 6</td>
<td>science major</td>
<td>2.90(0.21)</td>
<td>0.85</td>
</tr>
<tr>
<td>concept 6</td>
<td>non-science major</td>
<td>2.75(0.46)</td>
<td></td>
</tr>
</tbody>
</table>

*: p< .05.
**: p< .01.

Readers may doubt that the difference of each item’s difficulty level and the variability of content knowledge could affect students’ performance. In order to avoid ambiguity, an attempt was made in designing the study to control the similarity of difficulty level and content knowledge of the assessment items. Three of the five items with similar difficulty level relating to gas laws are randomly assigned as items 2, 3, and 6. Item 2 was assessed in the beginning while item 6 was measured at the end of the study. The statistical analysis of pair-wised t test was conducted to compare students’ performance on the same content knowledge of gas laws. It can be seen from Table 2 that both of students’ argumentation ability and conceptual understanding significantly improved (p< .001).

Table 2. Pair-wised t test result

<table>
<thead>
<tr>
<th>Assessment</th>
<th>pre-test mean (sd)</th>
<th>post-test mean (sd)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argumentation</td>
<td>2.50(1.15)</td>
<td>3.88(0.33)</td>
<td>5.22***</td>
</tr>
<tr>
<td>Conceptual understanding</td>
<td>1.94(0.68)</td>
<td>2.82(0.35)</td>
<td>5.00***</td>
</tr>
</tbody>
</table>
Did both science major and non-science major students benefit from the reflective peer assessment over the duration of the study?

In order to check the equity issue of educational opportunity and learning environment, we investigated the progress pattern and gap difference between science major and non-science major students' performance of argumentation and conceptual understanding. The science major students (n=15) were from colleges of science, engineering, and marine science while the non-science major students (n=15) were from colleges of humanity and art, management, and social science. It can be seen from Table 3 that both groups made progress not only on argumentation ability but also on conceptual understanding. The gap of argumentation ability between the two groups starts from a mean difference of 0.70 (2.93 vs. 2.23) in item 2 to a mean difference of 0.03 (3.85 vs. 3.88) in item 6. On the other hand, at the beginning of the asynchronous discussion the conceptual understanding mean scores of the two groups (2.27 vs. 1.68) were significantly different (p<.05) for item 2 and for item 3 (p<.01), but the difference gets smaller and gradually disappears week by week. There is no significant difference between the two group students' conceptual understanding starting from item 4 to item 6.

Discussion and implication for education

Previous studies found that problem-based learning in an asynchronous discussion environment were not successful for courses in physics (Kortemeyer, 2006) and in statistics (Hong, Lai, & Holton, 2003). Contrarily, in this study, we found that the participants made significant progress (p < 0.001) on their conceptual understanding and application in open-ended problem solving items. Possible reasons for the progress may be attributed to the learning opportunities gained by analyzing the quality level of the online answer statements and reflecting on other students' posted comments. Through the constant practices of these high level cognitive processes (e.g., analyzing and reflecting) and the continuous involvement in a specific learning topic such as gas laws, students were exposed to a learning environment that enabled them to formulate their own distinct opinion when they receive other students' rebuttals and criticism. Students integrated plausible content knowledge while contemplating different explanations of others' and finally constructing their own conceptual understanding and ways of applying these concepts in solving the next test item containing similar content knowledge but in a different context. In typical face-to-face classroom teaching, students are rarely afforded a long time to check and reflect on their own conceptual understanding and application of knowledge when solving problems. It is even harder to compare their own ideas with the ideas of others. The above finding led us to propose the approach of using reflective peer assessment in an online asynchronous environment for students to explore their misconceptions or misunderstandings and further construct understanding of scientific concepts. It is no surprise to us to find that the students made progress on their argumentation ability, since they were taught how to construct persuasive arguments. However, we are impressed by the improvement of the students' conceptual understanding and problem solving ability in the test items relating to gas laws—considering that they were not given further tutorial on the concepts related to gas laws in the study other than their high school physical science course.

Although additional studies are needed to confirm the practical utility, the initial finding of the diminished difference of conceptual understanding between science majors and non-science majors provides an indicator that the approach of reflective peer assessment has the potential to support student learning, particularly for those who have greater room for improvement (e.g., low achieving or non-science major students). However, this is not to say that this approach is not beneficial to high achieving or capable students. This study also found that students' majoring in science made progress in their conceptual understanding, as illustrated in Table 3. In total, the reflective peer assessment approach is likely to make positive contribution toward equitable distribution in learning outcome, and not at the expense of capable students. We suspect that the use of reflective peer assessment in asynchronous discussion provides opportunities for students to identify and discuss their alternative conceptions explicitly and publicly, which is helpful and constructive for students' conceptual understanding (Eryilmaz, 2002). This learning opportunity is rarely seen in traditional classroom teaching. In addition, when students work together collaboratively as a small group to answer the test items, it provides a working environment for them with less pressure than individual written tests. As Frenzel et al. (2007) indicated from their study of 1623 students, there was a close relationship between environmental variables and students’ emotional experiences. Furthermore, higher learning achievement was related to higher enjoyment and lower anxiety. If the conclusion of the above literature is persuasive, then educators and teachers are strongly encouraged to provide a learning environment that allows
students to publicly and explicitly discuss their understanding in their own words without any unnecessary pressure. Meanwhile, opportunities should be provided to students to work in interactive, cooperative, and collaborative teams.

The initial findings of the study shed additional light on the potential benefit of asynchronous discussions in promoting the development of high level cognitive abilities. In using the term “reflective peer assessment,” we intend to highlight the importance of providing opportunities for students to practice “reflective evaluation” and “evaluative reflection” in which the instructor assigns the roles as De Wever et al. (2008) recommends to students who take turns to serve as “answer provider” or “reflective evaluator.” The answer providers are encouraged to work collaboratively within their groups to reach a consensual conclusion in an assessment item with competing theories. In order to construct a persuasive conclusion, the students have to provide data and evidence and use warrants and backings to support their conclusions based on Toulmin’s model (Toulmin, 1958) that was introduced to them in the beginning of the class. In addition, each evaluator is responsible to exercise reflective evaluation, assign a score, and provide personal comments to the posted answer using the scoring scheme of Osborne et al. (2004). With the practice of assigning a score and writing comments, each student is exposed to the learning environment of practicing “reflective evaluation” (i.e., reflecting on the answer and then executing the evaluation). They are expected to learn how to provide counterarguments or rebuttals that disagree with existing arguments, or learn how to explain why they support a certain conclusion. Meanwhile, the asynchronous learning environment provides opportunities for students to review and reflect on other evaluators’ comments. Being evaluators, they are allowed to refine their comments and the original scores they assigned to the answer provider. In this stage, they are encouraged to practice “evaluative reflection” (i.e., evaluating other students’ comments and reflecting on their own comments). The evaluative reflection encourages students to observe others’ comments and critics. Based on McKendree’s (1998) assertion, observing a dialogue is beneficial for learning, especially when it is combined with reflection.

Despite the fact that the initial finding of the study is impressive and encouraging, readers are reminded that the sample size is relatively small. Therefore, caution should be taken in inferences of its quantitative results. In addition, care must be taken by making inferences from the research design of one group pretest and post-test which is not a true experimental design. In this study, since the participants have learned the content of gas laws in their high school years, no further gas law concepts were taught. The treatment was mainly used to help students clarify alternative conceptions and apply appropriate scientific concepts in contextual problem-solving situations through reflective analysis and criticism of peer answers and argument. During this period of time, the participants were not likely to have other learning resources in the specific topic of gas laws except the treatment. Therefore, the major potential threats of internal and external validity of this design can be reasonably avoided by selecting a learning topic where students (even some science teachers) have deep-rooted alternative conceptions (i.e., resistant to conceptual change) (Authors, 2000). Additionally, a longer period of interactive and collaborative dialogical reflections and argumentation would allow students to explicitly discuss and find the conflict between their own alternative conceptions and scientific argument. Further research studies focusing on different topics or subject matters of content knowledge with bigger sample sizes are strongly recommended.

References


Appendix: Sample item of the assessment instrument

As shown in the following figure, an empty flask is sealed with a rubber stopper which includes a glass tube. At the end of the glass tube, there is a drop of mercury. When the flask is immersed in a beaker filled with water of 3°C, the mercury will move to left. On the other hand, when the flask is immersed into a beaker filled with water of 80°C, the mercury moves to the right. If the whole set in the figure (not including the beaker) is moved from an indoor temperature of 25°C to an outdoor one of 5°C, can you predict and explain the movement of the mercury?
Supporting Mobile Collaborative Activities through Scaffolded Flexible Grouping

Ivica Boticki, Chee-Kit Looi1 and Lung-Hsiang Wong1
University of Zagreb, Croatia // 1National Institute of Education, Singapore // ivica.boticki@fer.hr // cheekit.looi@nie.edu.sg, lunghsiang.wong@nie.edu.sg

ABSTRACT
Within the field of Mobile Computer-Supported Collaborative Learning (mCSCL), we are interested in exploring the space of collaborative activities that enable students to practice communication, negotiation and decision-making skills. Collaboration is via learning activities that circumvent the constraints of fixed seating or locations of students. This paper presents one such collaborative learning activity that involves young students forming groups to achieve a specific goal (get the fractions on their handhelds to sum to one). Collaborative scaffolding is provided by the designed mobile collaborative application as well as by the students’ social relationships and the teacher’s facilitation. We report on our initial trials which show that the socio-technical design of the activity helps students in identifying their own strategies and stimulates collaboration. Beyond this specific application, we propose a generic model for mobile computer supported collaborative activities that can support a range of other tasks in learning languages, science or other disciplines.

Keywords
mCSCL, collaborative scaffolding, mobile learning

Introduction
The field of mobile computer supported collaborative learning has emerged in recent years spawning numerous technological designs for learning (Liu & Kao, 2007; Yin, Ogata, & Yano, 2007; Zurita & Nussbaum, 2004). Regardless of many contemporary mobile learning attempts focusing on out-of-class and contextualized learning, such as science centre visits, museum visits, field trips etc., (W. Chen, Tan, Looi, Zhang, & Seow, 2008; Y. S. Chen, Kao, Yu, & Sheu, 2004; Cupic & Mihajlovic, 2010; Fertalj, Hoic-Bozic, & Jerković, 2010; Jurcevic, Hegedus, & Golub, 2010; Kennedy & Levy, 2008; Klopfer & Squire, 2008; O’Malley & al., 2004; Rogers & al., 2002; Sharples, Lonsdale, Meek, Rudman, & Vavoula, 2007), we further investigate the potential of collaborative mobile technologies supporting collaboration in small groups (Colella, 2000; Dufresne, Gerace, Leonard, Mestre, & Wenk, 1996; Nussbaum, et al., 2009) in classrooms. As we want to promote collaborative learning amongst the students, we sought about designing collaborative activities in science and mathematics using these smartphone computers (Chan, et al., 2006; Looi, et al., 2010).

This paper presents a design for mobile computer supported collaborative learning in primary school classrooms in Singapore. Our research context is that we have been doing a two-year longitudinal design research study with a primary school in Singapore. We co-designed a whole year’s worth of lessons in science which are delivered through handhelds, specifically smartphone computers, and enacted these lessons over the course of one year. As such, the students in our experimental class are familiar with using their handhelds.

Our three-year study initially focused on science which inspired us to design a collaborative activity for math learning. In this activity, after students are assigned a fraction on their handheld, they are asked to form a group with other students in which the sum of their fractions is one. The activity completes only if every student belongs to a group with such a solution, thus requiring students to collaborate and to avoid or get out of the preferred social arrangements in order to help their peers in completing the task. During this study, we do not only rely on rich technological infrastructure (modern HTC Tytn II mobile phones and reliable 3G broadband internet connection). It is the classroom culture of 1:1 handheld device per student (Chan, et al., 2006) which allows us to critically read the affordances of our design from the student trial runs.

Building on this prior work in mCSCL, we explore the design of in-class mobile collaborative synchronous learning with flexible, small group sizes. We want to explore the space of collaborative activities in which students have to search and form their own groups in doing the activity. Such a socio-technical design of our collaborative activity is
intended to help students in identifying their own strategies of achieving the both the local and global goals via collaborative work.

This paper is organized as follows: in the next section, we provide a brief overview of recent developments in mCSCL in classroom settings. We will then present the design of our mobile-computer supported learning fractions activity. The subsequent section reports on our initial trials that seek to find out whether and how the collaborative scaffolding helps students in achieving both the local and global goals. In the final section, we propose a generic model for mobile computer supported collaborative activities that support a range of other tasks in learning languages, science or other disciplines.

Mobile Computer Supported Collaborative Learning

Early research in computer supported collaborative learning (CSCL) tends to foreground the role of computers as the focus of attention. Typically, each student uses a fixed-location glued-to-the-desk computer as the tool for collaboration. However, both the focus on the tool and the lack of collaboration actually happening have led to some skepticism in initial CSCL trials. It was felt that social interaction does not simply happen with a computer-based environment, thus emphasizing social and psychological dimension of the desired social interaction (Kreijns, 2003).

In advocating their approach to future classrooms organized around WILD (Wireless Internet Learning Devices), Roschelle and Pea argue that CSCL should leverage on application-level affordances such as augmenting physical spaces, leveraging topological spaces, aggregating coherently across all students as well as on the physical affordances of mobile devices (Roschelle & Pea, 2002).

mCSCL can be considered as a specialization of the field of CSCL. It alleviates the constraint posed by fixed times and locations for doing the collaboration activities. By employing mobile devices, learning becomes personal and mobile, and students are able to participate in collaborative learning activities when and where they want to (Looi, et al., 2010). Some research studies have shown that the use of mobile devices in classrooms could significantly impact student collaboration (Tseng, Hwang, & Chan, 2005). Students leverage on their own mobility and the mobility of the devices in order to coordinate collaboration and to exchange information simultaneously over the wirelessly connected devices.

One important research tackles the use of mobile connected devices in classrooms for the education of children of age six to seven (Zurita & Nussbaum, 2004). These students were given language and mathematics tasks they had to solve by working in groups. In the process, they had to exhibit a certain level of interaction and communication in order to complete the group tasks. The authors report that the use of wireless networks in the classroom opened up many educational possibilities and that mobile devices advance various components of collaborative learning, namely, the learning material organisation, social negotiation space, communication between team members, coordination between activity states and the possibilities for interactivity and mobility of team members (Kreijns, Kirschner, & Jochems, 2002). Concerning the main advantages of mobile versus classical computer supported collaborative learning, enhanced possibility for communication, negotiation and mobility has been proposed (Zurita & Nussbaum, 2004). Together with appropriate design of learning activities, a network infrastructure of mobile devices can support collaborative activities in which students extend their area of communication and mutual interaction.

In their conceptual framework for mCSCL (Zurita & Nussbaum, 2007), the authors take an activity theory approach by building on the Engestrom’s expanded Activity Theory (AT) model (Engestrom, 1999) and identify three main components of the mCSCL activity system: the Network component, the Rules and Roles component and the Collaborative Activity component spanning across so called social and technological activity dimensions.

Grouping criteria have also been claimed to have impact on mobile collaborative learning. In a study on the impact of grouping criteria on socio-motivational aspect commonly used to evaluate collaborative learning (Zurita, Nussbaum, & Salinas, 2005), authors have determined that “when the children select their group mates (so called Preference criterion) more social behavior aspects with significant improvement can be observed” (p.159). The study therefore foregrounds personal students’ preference towards their classmates as the top grouping criterion in order to achieve a more collaborative environment.
We are interested in exploring and designing the space of collaborative activities which enable students to practice communication, negotiation and coordination skills in the process of forming their own groups to solve a group goal. In our approach we supplement the two-level (social and technological) network analysis with a spatial network. The spatial network allows us to more precisely pinpoint the social process in our pursuit of analyzing learning and collaboration occurrences. By employing flexible grouping approach, we allow students to choose their own groups depending on their personal preferences. Since there is a reported negative effect of personal preference grouping criterion on negotiation (Zurita, et al., 2005), we introduce more structure to the activity by using technological scaffolding in order to channel student grouping choices.

**Design of FAO collaborative application**

In this in-class activity each student has a handheld device. Once they launch FAO, their handhelds are connected to FAO server through a 3G wireless network. Fractions are depicted on students’ mobile devices in form of circle sectors (slices) (Figure 1). Students have to collaborate in order to merge (add) fractions. They have to identify peers with complementary fractions (with respect to getting a sum of 1) and then invite them to form groups (Figure 2). The main goal of the assignment for each emerging group is to form a full circle (a whole) by combining circle sectors (graphical representations of fractions). Inter-group collaboration and negotiation may be necessary to complete the task.

![Figure 1. A fraction assigned and displayed on a student’s mobile device](image1)

![Figure 2. Student issuing a group invitation to his classmate](image2)

Collaborative scaffolding is provided by the designed mobile collaborative technology, students’ existing personal relationships and the teacher’s facilitation. We analyze student participation in the activity through three networks: technological, social and the spatial network which enables a form of embodied participation and is formed by the dynamical rearrangement of the students as they move about with their devices. The three networks together provide the infrastructure for supporting coordination, communication, negotiation and mobility.

Based on the screen information available on their handhelds, each student can access her own fraction as well as access the fractions of all other students in the class. This provides the technological level of support for the activity. The students also rely on their social network of close friends in the class. They are more likely to invite their own friends or their own gender friends to form their own group which provides the social level of support. As the students are mobile, they re-arrange their spatial configuration as they move. It is also likely that they interact with those who are near them spatially.

**Phase I of activity: Distribution of Fractions**

As soon as all the students have turned their devices on and the teacher started the activity, the server registers the total number of students and then runs an algorithm to randomly assign a fraction to each student (Figure 3). The
algorithm ensures that there is a global solution, namely, a configuration of groups of students in which every student belongs to a group and every group completes its task. Although the random fraction distribution ensures fraction diversity, the teachers can control the type of fractions distributed therefore structuring and fine-tuning the activity.

In Figure 4, the system detected five students as potential activity participants and assigned them with randomly generated fractions. The generated fractions are 1/2, 1/2, 1/2, 1/3 and 1/6 and are displayed on students’ mobile devices. In this first phase of the activity, students ponder about their individual fractions and try to find out what are the other generated fractions in order to figure out the possible ways of forming groups.

**Phase II of activity: Negotiation and exchange**

To identify the potential candidates in order to form a group, a student can rely on the graphical user interface of her mobile device and browse through the list of all available students and their fractions (Figure 5) or they approach the problem through face-to-face interactions and detect the potential candidates through conversation. When a student identifies another student with whom she could form a group, she uses her mobile device to issue the group invitation. The request is then dispatched to the server side which forwards it to the invited student (Figure 6). Through a series of invitations, accepted and rejected requests, students arrange themselves to form groups.

Some students may have some difficulties with adding up the fractions or with reaching some local optimum (Figure 7). Local optimum presents a formed whole circle within a group. Although optimal for a group, it might not be optimal for all groups. Some groups might be blocked in reaching their local optimal solutions because one group reached a certain local optimum. The group then has to be broken and other groups have to be assembled, hopefully leading to optimal solutions for all groups which leads to the completed activity.
Phase III of activity: Towards the global-oriented goal of all FAO

Figure 7 shows an activity with six students who have already been assigned with fractions. Through the collaboration they were able to form two groups but are unable to accept the one leftover student (holding 1/2) since adding his fraction would make both group fractions larger than a whole. In order to achieve both the local and global goals, students have to negotiate and to re-form their group memberships. Thus, in addition to the individual goal of forming a whole, students have to work collaboratively in order to achieve the common goal of all groups having a full circle. Nevertheless, while some groups might have formed their wholes (their individual collaborative goal is achieved), the others might have reached a dead-end situation, and be unable to proceed. This is a situation where students are required to put the global goal before the individual or group goals and to try thinking collaboratively about other possible solutions or group configurations. Only when each group has formed a whole is the activity over.

Trials of the Collaborative Activities

The proposed design was evaluated through a series of trials with the primary school children roughly aged 8-9 grouped in groups of 8 and 16 students divided in two batches. The first batch of students was introduced to the software and the “ways of doing the collaboration”. Students had some prior experience in using different mobile learning tools and needed just a brief overview of the FAO software. The second batch of students was not familiar with the “ways of doing the collaboration” such as how to invite other students to their groups, how to negotiate for their cause, align their personal goals with the overall group goal etc.

In order for the trials to mimic the actual classroom arrangement, the research team worked closely with the teachers of two classes. All students received the instructions on “how to do the collaboration”. This included simple rules such as: “when you are in the same group stand together”, “you are allowed to talk to other students in addition to working on the device”, “do not automatically reject group invitations, and talk to your colleagues to see their needs” etc. Concerning the students’ knowledge of fractions and fractions operations, it is important to note that understanding fractions do present a challenge to some of the students as one student critically commented: “Fractions are worse!” meaning the most difficult.

How Groups Emerged in one Trial Run

We look into how students perform in one trial run. We use a visual coding scheme which shows the spatial distribution of the students who are identified by their individual fractions, their current grouping, and their gender dimensions. Male students are shaded and named with abbreviations starting with M, while the female students have the names starting with F. Their position and mutual distance in the picture reflects actual position and distance taken during the game. In the beginning of the activity, the students started exchanging ideas about arranging fractions (denoted by the two-direction arrows) (Figure 8). The discussion started to grow from pairs to groups of three and four students (Figure 9).

As the activity progressed, two groups were almost simultaneously formed indicating positive outcome of the negotiation activities. Following the successful creation of two groups the third group was created (Figure 10).
Although the system provided the student with the flexibility of choices of choosing other students (M1, M3, M4, F1, F2, F4 could all make pairs with each other), personal and gender preferences influenced the way groups were formed. This had an impact on the dynamics and complexity of the activity: as it progressed, the overall number of the possible combination decreases making the choice of partners more straightforward.

Two students (F3 and M2) were applying the “combining same fractions strategy” and were not aware them joining could lead to a whole. They decided to seek peers’ assistance in identifying the possible solution for the activity (Figure 11). Not able to independently make the decision, student F2 was dispatched to seek the assistance from the teacher. In the meantime, the discussion between other team members continued (Figure 12).

After some additional consultation with the other teams and some teacher facilitation, students F3 and M2 finally managed to form a group leading the overall team effort towards the end (Figure 13). Figure 14 illustrates the spatial group arrangement for all students while the Figure 15 provides a close-up view of the intra-group interaction during the activity.
Negotiating Local and Global Goals through Backtracking

The previous trial run is straightforward in the sense students did not have to backtrack, meaning there was no need to disassemble the groups they are in and to assemble new ones. The focus was on achieving individual goals and yet at the same time, the global goal is reached. There are other runs in which students get into groups which require them to disband and re-group, and enter into new negotiations leading to new group configurations in which every student belongs to a group which achieves the goal of having the sum of their fractions as one.

We discuss such a case in a trial run with 14 students using a series of screenshots from the teacher’s console showing assembled groups. Figure 16 shows the 8th step (the 8th group configuration) in the activity. Students negotiated their way to this step in a fairly straightforward fashion: they employed a simple strategy of combining the same fractions to create the group (1/2 went together with 1/2, 2/4 with 2/4, 1/3 with 1/3 and 1/6 with 1/6). When several choices of students with the same fractions become possible, the students employed the strategies of personal and spatial preference in choosing their partners.

![Figure 16. 8th group configuration in the FAO activity](image1)

![Figure 17. 9th group configuration in the FAO activity (after the change in the strategy)](image2)

![Figure 18. Graph of number of groups in the activity through time](image3)
After the 8th step, the students were faced with the consequences of the chosen strategy: they were not able to proceed by simply combining the same fractions because of a simple reason: there were not enough fractions to combine. On the other hand, there were other fractions they could use but their strategy had to be changed or supplemented. After a short period of discussion, social interaction, peer instruction and some scaffolding provided by the technology and the teacher, the students were able to advance to the next step of the activity. The teacher scaffolding in this case consisted of a teacher pausing the activity, explaining the current group configuration to students, explaining students they have break their current groups in order for the activity to proceed and helping them find the potential peers on their own (Figure 17).

As the activity progressed, the number of groups could only drop or remain the same. To identify the deadlocks due to the backtracking, the teacher tracked the number of active groups on her computer (Figure 18). If the number of groups remained constant throughout a longer period of time as shown in Figure 18, the teacher could intervene and provide some additional scaffolding by advising some groups to re-group.

**Strategies adopted by students**

The activity began with researchers giving some brief instructions to students on how to use the device and the software. The series of trials show that the children have no problems with the use of the technology mostly because they have had the chance to use handhelds (HTC Tytn II devices) throughout a prior series of lessons. They were both familiar with the pre-installed software solutions such as Excel and Word and were taught how to use custom-made software solutions for mobile learning. Therefore, there were no usability issues, and the students were able to quickly familiarize themselves with the user-interface and used the application effectively.

After the initial period of confusion, resulting in a short period of silence throughout which the problem of forming wholes out of fractions was mentally processed, the activity continued. In addition to individual effort of examining the list of all students and their assigned fractions, some strategies started to emerge and were shared amongst the participating students. The main strategies are: looking for the same fractions in order to form a whole, identifying students to invite to be in one’s group based on what other fractions are needed to make a whole, gender and personal preferences, and just randomly sending out invitations.

Prior to distributing the fractions to students, the system generated the fractions with a specially designed algorithm designed to achieve two main goals: diverse randomized fractions and achievable final solutions (the global goal or solution). This means that some students received the same fractions (e.g. 1/2 and 1/2) They started looking for peers with the same fractions in order to create groups which in some cases turned out to be a successful strategy, while in other it caused impasse situations and required some backtracking prior to achieving the group goal. Here are some conversations of the students while exercising “choose the same fraction” strategy:

- One student explained her self-employed strategy: “If I have 2/4 I go and look for other 2/4”.
- One student asked out loud: “Who does not have a fraction?” He then approached a student without a group: “How much do you need?” Another student joined the discussion: “You have 1/2!, you need 1/2”. The student without a group responded: “But it is only Kenny [another student a bit further away] who has 1/2 and every time I invite he says do not want to.” The second student suggested the way for the student without a group to proceed: “Then just talk to him.”
- One student took on a mediating role and circled around the room trying to identify who should join with whom. After a while he suggested out loud: “Clifford and Wendy both have 1/2”. Then he spoke directly to Wendy: “You will have to go with Clifford”.

The technological layer provided some scaffolding during the process: the students had a list of all other students with their assigned fractions. In addition to the individual approach of identifying the same fractions from the list, the students could switch to the social and spatial network to receive some additional technological scaffolding from comparing concrete and abstract representations of their fractions.

The students had many different configurations to choose from when assembling the groups in order to reach the global goal. Most of them utilized the spatial network to approach physically nearest peers and try to make a group together with them. Since students’ personal and gender preferences controlled initial spatial activity arrangements, students were able to take some ownership of the activity.
One impasse faced by the students occurred when the software did not allow them to increase the sum of their fractions beyond one. These students were surprised by the system message of being unable to allow the group of just two members (e.g. members with 2/3 and 1/2). To get out of this impasse, they had to question or relook at their strategy of merging any two students and looking for the third member to complete the group.

Almost all students approached the activity only with the individual goals in their mind without thinking about the global goal. They had to be reminded on numerous occasions about it and were encouraged by the teacher to assist their peers in rearranging or perhaps even breaking their own groups. The understanding of the shared goals was perhaps the most difficult for the primary school children to grasp. Nevertheless, some students acted as mediators, being able to cover both the task and the ways of connecting individually oriented students.

**Learning to collaborate**

Learning how to collaborate proved to be another demanding task for the primary school students. Achieving the local and the global task goals required them to extend their social circles and go beyond their social comfort zones. The activities started with fixed socio-spatial arrangements: girls standing in line with girls and boys co-located with other boys. In order for the activity to progress, students had to exit these configurations: one of the first identified endeavours was “crossing to the other side” in order to negotiate a new group formation.

An interesting case emerged with a boy and a girl not able to collaborate even though the overall activity progress depended on it. Since the face-to-face negotiation was out of question for them, they relied on the technological layer to send out group invitations. It seems the technological medium facilitated to help them overcome their pre-established personal preferences.

In some cases in which the technological support was not adequate, social scaffolding came into play. Students encouraged each other to form new groups both verbally (giving explanations on why to go to another group) and physically (gently pushing their peers towards the potential partner). At times, social collaborative scaffolding was powerful enough to overcome personal preferences for group membership.

In contrast to overcoming personal preferences in achieving both individual and group goals, some students built on top of their personal relationships and spontaneously offered help to their colleagues. After a group of two girls was created based on personal preferences, they together decided one of them should accept a new group invitation. After their group was dismantled, the girl left alone was offered some help in identifying her new mates.

In the process of dealing with impasses and employing and testing the new strategies, the students provided peer instruction to help each other. For example, they had to convince their peers to adopt new strategies. One student identified that two fractions (2/4 and 1/2) can be merged in order to form a whole. In order to convince his colleague, he used this simple explanation strategy: “You have to increase your fraction!” Although the choice of words was not appropriate (one might understand it as to search for a larger fraction/number), the student later clarified his advice by pointing out that 1/2 equals 2/4.

**Towards a Generic Model for Collaborative Scaffolding in mCSCL**

Through the conducted trials several sources of collaborative scaffolding were identified: technological, teacher and social scaffolding. All the three components are the sources for collaborative rules which structure student participation in the activity both in the sense of social interactions and task completion. Technological scaffolding provides technology-embedded structures or rules for sending and receiving messages through the handhelds. It relies on a specific rule structure and their interconnection, and is triggered via the user interfaces transmitting the messages. Social scaffolding, on the other hand, builds on top of collaborative rules predefined by the teacher but draws from the emergent collaborative practices such as peer instruction, sharing through discourse, and mediation. The teacher scaffolding provides contextual assistance supplementing both technological and social scaffolding but mainly builds on top of the existing individual and collective group competence. The teacher scaffolding consists of teachers stepping into the activity at critical points (e.g. students cannot move from one group configuration to the other) in order to facilitate the activity progress. The teacher typically starts a discussion about the problem students
have, and try leading them to a possible solution. In the process, teachers can combine technological and social scaffolding thereby delegating some work to the technological infrastructure or the students (Figure 19).

![Collaborative scaffolding](image)

*Figure 19. Social, teacher and technological scaffolding as the main elements of collaborative scaffolding*

The collaborative scaffolding can be applied to different learning content. Besides activities for learning fractions, collaborative activities can take the form of composing sentences, or forming Chinese characters or idioms by using the same set of social and technological collaborative rules. In the software design, the rules and logic for a particular domain are specified in the Generic Content Rules and Logic interface (Figure 20). The system looks at any mobile learning content as the sequence of content elements that can be combined in a sensible unit, and distributes the elements (either generated automatically or as provided by the teacher) to students.

The content specific rules are separately defined for each mobile learning application. The fractions activity of FAO comes with rules which define answers to questions such as: How to combine fractions (by summing or some other operations)? What makes a whole or a solution? How to generate fractions prior to distributing them in order to have feasible local and global group goals? How to introduce complexity when generating fractions (such as having larger denominators)?

With a collaborative activity for the composing sentences, the basic content elements assigned to and manipulated by the students could be words and phrases. The rules and logic would need to deal with the following issues: How to combine words or phrases to form sentences? How to obtain or generate words and phrases prior to distributing them to students? How to check the validity of a constructed sentence in case there are more feasible solutions than the teacher predicted?

With a collaborative activity for forming Chinese characters, the basic content elements are radicals which are arranged spatially in correct ways to form legitimate Chinese characters. The rules and logic would need to deal with the following issues: What are different graphical layouts of Chinese characters? How to check whether a combination of Chinese characters produces a valid character? How to check the semantics in case there are more feasible solutions than the teacher predicted? Figure 20 shows our model for designing a generic collaborative software for supporting the design of different collaborative activities.

A feature is required for the teacher to be able to specify activity parameters that directly impact the complexity of the activity, and the possibilities for collaboration in the activity. For FAO, the teacher provides parameters which determine whether the visual pie-chart representation should be displayed alongside with the mathematical notation of fraction, which denominators are allowed to appear in the game, whether to show fractions subdivision etc. All of these determine the difficulty level of the activity and the level of scaffolding the students receive from the technology. In the sentence composition activity, the teacher inputs the text and specifies the way it should be distributed to the students. The teacher can choose to distribute a word per student or to decompose sentences into phrases. In the Chinese character activity, the teacher defines the set of Chinese characters or radicals to be made available to students, and therefore indirectly determines the range of different characters that can be composed.
It is not only the use of different content that makes the system generic. Collaboration rules that utilize the content are generic as well, allowing users to collaborate around different content bits (e.g. fractions vs. Chinese letters). Generic are the communication mechanisms as well, allowing the transfer of messages aimed at different content areas.

**Conclusion**

The paper presented the design of a collaborative activity of learning fractions with handheld computers and the findings of some preliminary trials. Primary three students used handheld devices and specially designed software to participate in a collaborative effort of achieving local goal of forming groups with wholes and a common global group goal. The activity as supported with collaborative scaffolding consists of three main scaffolding sources: technological, social and the teacher. Technology provides scaffolding in the sense of both generic and context-specific rules and logic, while the teacher acts as facilitator and helped the students in dealing with impasses. Social scaffolding is encouraged in order to increase student interaction and collaboration.

In our trials, students were able to come up with some ad-hoc strategies of doing the activity and solving the problem, some of which inevitably ended with impasses which had to be resolved with collaborative scaffolding. Students were able to modify their initially chosen strategies and realized the importance of achieving the global goal besides their group goal, therefore learning how to collaborate with these interdependencies.
We feel it was the interplay of technological, teacher and social scaffolding which contributed to the overall progress of the activity. The technological and social scaffolding were interchangeable depending on the personal preferences of the students. Students armed with good communication and negotiation skills relied more on the social scaffolding, while more introverted students used the device as a medium of carrying out actions that would otherwise probably never be externalized. It was the technological scaffolding that made the activity progress easier further advancing student problem solving skills. Instead of personally checking other students’ fractions, students could refer to their devices and browse the corresponding lists. Furthermore, students were able to switch between different kinds of problem presentation and see their group artifact at anytime. The teacher scaffolding when introduced at critical moments bridged the gaps neither technological nor social scaffolding could therefore preserving the momentum of the activity.

Building on this specific application of a fractions activity, we propose a generic model for collaborative scaffolding in mCSCL that enables the design of collaborative learning scenarios for handheld computers in different domains such as sentence or character construction in language learning. The characteristics of our collaborative activities include interdependency on other students to form a group solution, agency in students to accept or reject invitations to join groups, reliance on collaborative skills to find collaborative partners, emergent groups instead of fixed groups, facing the tension between meeting a group goal vs. meeting the global goal, and willingness to backtrack group solutions in order to seek a global solution.

In our trials, we faced a host of issues ranging from classroom management to technical glitches. In the trials with large groups (whole classes of 40), the students typically exhibited the strategy of randomly sending out invitations therefore checking each other's progress (they had to wait a long time for each other’s reply). In one particular case, as the waiting time between the steps in the activity went too long, the progress of the activity was disrupted. This leads us to a new cycle of system and user interface re-design in which students do not rely so much on the initially chosen request-response design paradigm, but rather choose from a set of available group configuration publicly displayed on the common shared screen.

In our approach, we have chosen the 3G network connection as the means of exchanging system messages triggered by users. We advocate this approach due to several factors: connection stability, signal coverage and the decreasing cost of such network connectivity. It is our belief that this presents a significant advantage over the free WiFi connectivity and opens up possibilities for activity and system extensions to outside of classroom boundaries.

In addition to the fractions activity presented in this paper, we have conducted a series of trials including Chinese character learning and plan for another full-fledged semester-long study dealing with the issues of regular lecture integration. With the redesigned technology and a slightly adjusted research design, we hope to more thoroughly explore the impact of this technological innovation in regular classroom environments.

Acknowledgements

This paper is based on work supported by the Learning Sciences Lab and a grant from the National Research Foundation, Singapore (Grant #: NRF2007IDM-IDM005-021).

References


QTIMaps: A Model to Enable Web Maps in Assessment

Toni Navarrete, Patricia Santos, Davinia Hernández-Leo and Josep Blat
ICT Department, Universitat Pompeu Fabra, C/Roc Boronat 138, 08018 Barcelona, Spain // toni.navarrete@upf.edu // patricia.santos@upf.edu // davinia.hernandez@upf.edu // josep.blat@upf.edu

ABSTRACT
Test-based e-Assessment approaches are mostly focused on the assessment of knowledge and not on that of other skills, which could be supported by multimedia interactive services. This paper presents the QTIMaps model, which combines the IMS QTI standard with web maps services enabling the computational assessment of geographical skills. We introduce a reference implementation of the model, with Google Maps as the web map service, comprising both an editor and a runtime system, which have been used in two learning situations. The tooling developed and the real use results demonstrate that the QTIMaps model is usable and provides educational benefits. We describe three other assessment activities, showing that the model can be applied to a variety of educational scenarios.

Keywords
E-Assessment, geographical skills, QTI, interactions

Introduction
E-Assessment approaches exploit interactive technologies to support and enhance educational assessment (Conole & Warburton, 2005). Nowadays it is widely accepted that assessment processes, as integral part of education, should support a variety of skills including general and subject matter abilities (Bennet, 1998). The European Qualifications Framework define skill as ‘the ability to apply knowledge and use know-how to complete tasks and solve problems. Skills are described as cognitive (involving the use of logical, intuitive and creative thinking) or practical (involving manual dexterity and the use of methods, materials, tools and instruments).’ Based on Bloom’s Taxonomy (1956) there are six levels of skills complexity that go beyond knowledge, i.e., comprehension, application, analysis, synthesis and evaluation. We address the problem occurring when some skills are included in the educational curriculum but the methods to assess them are not appropriate (Bennet, 1993; Boyle, 2009).

This paper focuses on spatial or geographic information, its representation and related skills. Spatial thinking, one form of thinking, is a collection of cognitive skills, according to the committee of the American National Research Council that has analyzed the incorporation of Geographic Information Science across the K-12 curriculum. The skills consist of declarative and perceptual knowledge and some cognitive operations to transform, combine, or otherwise operate on this knowledge. The key to spatial thinking is a constructive amalgam of three elements: concepts of space, tools of representation, and processes of reasoning, and these elements are present in the K-12 curriculum of most countries. For example, the Spanish curriculum states that secondary education learners should be able to “identify, localize and analyze geographical elements at different scales” and “search, select, understand and relate (...) cartographic information from different sources: books, media, information technologies” (Spanish Government, 2006). Learners should also be capable of understanding processes and how humans influence the environment.

Maps have been the main representation of geographic information for centuries. Information technologies provide useful tools enhancing traditional maps in terms of interactivity and ease of modeling spatial processes. Geographic Information Systems (GIS) support the generation of maps from spatial information, but moreover its analysis (finding patterns in geo-referenced data, for instance) and modeling (such as how the environment will evolve). A GIS should be able to answer the five questions formulated by Rhind (1992), and reflected in Table 1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Type of task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is at ...?</td>
<td>Inventory</td>
</tr>
<tr>
<td>2. Where is ...?</td>
<td>Monitoring</td>
</tr>
<tr>
<td>3. What has changed since ...?</td>
<td>Inventory and monitoring</td>
</tr>
<tr>
<td>4. What spatial pattern exists ...?</td>
<td>Spatial analysis</td>
</tr>
<tr>
<td>5. What if ...?</td>
<td>Modeling</td>
</tr>
</tbody>
</table>

Table 1. The five types of questions a GIS should answer
Nevertheless, GIS’s are still mainly oriented to professionals and are difficult to use by the general public, and in particular by K-12 teachers and students, while web maps also enhance traditional maps providing more interaction possibilities. Google Maps is the most popular web map application, offering a world-wide cartography including street-level information for many countries as well as satellite/aerial images at different resolutions. While Yahoo Maps and Bing Maps, are products of Yahoo! and Microsoft with features similar to those of Google Maps, other web maps are compliant with international standards and can show information from Spatial Data Infrastructures (SDIs). An SDI is an infrastructure for sharing geographic information (and meta-information) in a standard way. SDIs are usually created and maintained by public organizations that publish a large amount of geographic information on different topics, such as environment, economy, demography or climatology that can be very valuable in education.

Apart from allowing to access this huge information, web maps enable the combination of different information sources in a single map, and allow the user to identify and draw elements on the map. From the five questions formulated by Rhind, web maps completely support monitoring and inventory, while analysis and modeling only partially. The rich interactions that web maps provide can be used in learning activities for the learner to understand and reflect on space. Web maps can also be used to create activities to assess the acquisition of skills included in the curriculum, both in a formative or summative way.

Very related to web maps is Google Earth (in fact some its functionalities have been recently integrated in Google Maps). Patterson (2007) demonstrates that Google Earth has a considerable potential to enhance teaching methods and supports the development of spatial thinking and other skills (such as critical analysis). Google Earth, as a desktop application, is difficult to integrate in web-based e-Assessment platforms, but is very related to web maps - in fact some its functionalities have been recently integrated in Google Maps. As the potentials described by Patterson can also be applied to web maps, we will focus on them.

ESRI (1998) states that around 80% of data in information systems have a spatial component and consequently can be portrayed in a map. This percentage can also be high in several subjects. Location has a significant role not only in geography, but also in other fields such as history, dialectology, biology (botany and zoology), geology or sociology, among others. SDI’s provide information about most of these topics, especially those related to the environment.

However, current learning and assessment e-environments cannot deal with web maps. In fact, this happens for other media with specific types of multimedia interaction. Video sequences composers, chemical molecules viewers/editors or history timelines are neither well supported by these e-environments.

Centering on e-Assessment, there are very few skill-oriented approaches related with maps, and are mostly closed in terms of technologies used and possibilities offered to teachers to customize the designs according to the educational needs (Ridgway & McCusker, 2003). Klaus & Leonhard (2004) propose an innovative assessment tool with their own implementation of interactive maps, but the tool offers a limited number of maps and question types which do not answer the requirements of specific learning situations. Chang et al. (2009) propose a map-based assessment application using Google Maps that allows teachers to edit four types of items (multiple choice, essay question, sequence and map-based questions), but users cannot interact with the maps to answer the questions.

A limitation of these approaches is that they are not using any educational technology standard. Standards should play an important role in e-Assessment tools because they: (1) provide a data model to computationally represent questions, items and results, and (2) enable content reuse across different learning environments. IMS Question and Test Interoperability (QTI) (IMS, 2006) is the e-Assessment de facto standard and interoperates with other IMS specifications such as IMS Learning Design (IMS, 2003). QTI supports questions involving graphical interactions, such as hot spot or select point, among others. Moreover, although QTI does not address interactions dealing with web maps, it supports the definition of new types of items adapted to specific educational goals. Consequently, a valid solution to the integration of web maps (and any other interactive media) in assessment scenarios can be based on the QTI standard.

In this paper we introduce the QTIMaps Model. QTIMaps adopts, extends and combines QTI with web maps services to provide solutions for the assessment of spatial thinking and related skills. The QTIMaps environments will both benefit from the open aspects of QTI and webmaps services: interoperability with other learning services.
and standards, and integration of large amounts of up-to-date geo-information provided by SDI’s. To validate the model, we present a QTIMaps implementation and its use in a secondary school class and an informal workshop. The implementation uses Google Maps as an example of a web maps service and encompasses an editor and a player based on a QTI engine. The evaluation results extracted from the use of the implementation show that the QTIMaps model is applicable for its intended use, both in terms of educational benefits and usability. Moreover, we illustrate with three additional scenarios that the QTIMaps model supports a variety of assessment activities.

Although QTIMaps is related to (Bouzo et al., 2007), a system for enhancing a QTI engine with maps from Google Maps, the current model is much more complex. QTIMaps model provides a more flexible way to define interactions based on the introduction of interaction methods and validations rules, as will be discussed below. Furthermore, QTIMaps is not restricted to a given implementation of web maps such as Google Maps, supporting a variety of layer types, including those published in SDI’s, that can be combined in the same map.

The first section of the paper summarizes the state of the art in the area of e-Assessment specifications and map services. Next, the QTIMaps model is explained. Then, we present the implementation of the model, including an editor and a player, followed by the situations in their real use. The following section introduces other scenarios to show the applicability of the model in formative and summative assessment activities. The final section is devoted to conclusions and to explain the main lines for future work.

State of the art

The QTIMaps model combines the use of a widely accepted assessment standard (QTI) and existing map services, whose state of the art is discussed next.

QTI and related tools

QTI (IMS, 2006) provides an interoperable data model for the representation of questions (items), tests, their outcomes and feedback. It enables thus the exchange of questions and item banks between QTI editors or composition tools, but also with learning systems which include assessment or its outcomes. The data model is defined in abstract terms, but an XML binding is also provided. The latest QTI versions are 2.0 and 2.1. QTI 2.0 focuses on the representation of individual questions categorized by their types of interaction, and contains a set of graphic interactions to deal with images. QTI 2.1 extends QTI 2.0 by considering the aggregation of questions in tests organized in sections, and defines sophisticated ways of producing outcomes reports for the whole test.

A QTI compliant editor (used by a teacher) will turn the test and questions into a set of XML files containing all their information. Similarly, the students’ interactions and answers using a QTI compliant player can be assessed automatically and feedback provided.

A QTI engine is the software component managing the QTI data model, processing the XML files and generating the outcomes according to the user actions. We should mention two open source QTI compliant engines: (1) The APIS QTI 2.0 engine (APIS, 2004), which is a modular item-rendering engine addressing the operations required by potential tools as defined in the Open Knowledge Initiative (OKI, 2001) and IMS Web Services; (2) the R2Q2 (Rendering and Response processing services for QTIV2 Questions) (Wills et al., 2006), (R2Q2, 2006), an implementation built from scratch, aimed at providing a complete renderer and response processing engine. Both implementations are limited to the QTI 2.0 (thus only able to process individual items) and cover only some of the question types defined in the data model.

In previous work we considerably extended the items types beyond those covered by APIS and built QTI 2.1 compliancy, resulting in what we called the newAPIS engine (Blat et al., 2007). NewAPIS functionalities include dealing with test context beyond just questions, a wide range of new elements at the test level, and more complex response processing.
Standards for geographic information

Serving maps on the Internet has become very popular in recent years, and a lot of mapping servers implementations exist, both commercial and open source. The Open Geospatial Consortium (OGC, 1994), an international organization composed of companies, governmental agencies and universities, defines consensus geospatial services standards; several OGC specifications have already become ISO standards.

OGC has developed the WMS specification, which defines a service-based interface for a standard web map. The OGC Web Feature Service (WFS) specification is widely used to publish geographic information in vector-based formats. These OGC standards are at the core of the Spatial Data Infrastructures.

Originally developed by Google, KML is a simple XML-based file format to store vector-based geographic information that has been widely popular and became eventually adopted by OGC. GML is another XML-based format standardized by OGC, not popular amongst the general public, but widely used by professionals to share data.

The QTIMaps Model

Web maps provide both new ways of presenting information and new interaction possibilities. QTIMaps is a conceptual model enabling the creation of new types of questions, which integrate the interaction possibilities of web maps (see Figure 1).
The model is based on the principle of separating the definition of multimedia interactions not supported by QTI from those already supported. This minimizes the changes required in QTI-compliant tools to support new interaction types. A class called `MultimediaInteraction` has been defined to represent any type of multimedia interaction not included in the standard. An instance of `MultimediaInteraction` can be associated to a QTI question (`AssessmentItem` - The MultimediaInteraction instance is related to the `ItemBody` of the `AssessmentItem`) to integrate the interaction into the question. Different types of media interactions would be modeled as subclasses of `MultimediaInteraction`. The `InteractiveMap` class is the subclass of `MultimediaInteraction` that represents a web map interaction.

**The InteractiveMap class**

To configure how an interactive map is shown, two main properties need to be defined: its spatial extent (the rectangle of the Earth shown in the map, i.e. the geographic location of the map) and its scale, which actually contains the scale denominator (in a 1:1000 scale). Other attributes may specify whether the user can freely pan or zoom or s/he is restricted to a given extent and set of scales. The spatial reference system -SRS- (srs property) of the map must be defined. The SRS (also known as spatial coordinate system) provides a framework for the map coordinates, depending on how the Earth is modeled to obtain the map. An SRS is identified by a code defined by the European Petroleum Survey Group (EPSG), whose list of SRS codes is a de-facto standard for the geographic information community. Finally, the `units` attribute specifies the units of measure of the map, such as decimal degrees, meters, miles, etc.

**Layers**

An interactive web map can contain one or more layers of information from one or different sources. Different subclasses of the `Layer` class are defined to deal with different types of layers. The simplest map contains one layer - from Google Maps (or Yahoo Maps or Bing Maps), showing the “satellite”, “map”, “hybrid” or “earth” view depending on the appearance attribute. The `variableAppearance` attribute determines whether the user is allowed to change the map appearance.

These types of layers are very useful for many assessment items, but a huge amount of geographic information is published through SDIs. The WMS and WFS classes provide support to incorporate these specifications compliant layers through which SDIs publish information: WMS which returns a raster layer and Web Feature Service which returns a vector-based one.

A lot of users and organizations publish their own geographic data through KML or GML, KML and GML classes are defined to deal with this data. As WFS, KML and GML deal with vector-based information, a style (class `Style`) can be specified to determine how the elements are symbolized in the layer, and therefore shown in the map/tests. For instance, a GML file may contain the geometry (polygon) of a natural park, while the style determines that it will be shown, as a red 2 pixels wide border line, its interior in orange with an Arial bold black label showing the name of the park. Styled Layer Descriptor (SLD) is the OGC style standard.

**Geographic Elements**

An question dealing with a web map usually requires the learner to identify or to draw geographic elements (such as cities, countries or natural parks) on it. These elements are represented in two ways, by means of either a marker (class `Marker`), or a geometry (class `Geometry`). A marker is an icon placed at a certain location, displaying a message when it is clicked. The marker properties are its `icon` and its `message`, which contains some HTML code. The geometry gives a more precise spatial location representation: it can be a point (similar to a marker), but also a complex line (with several segments) or a polygon, as well as their corresponding collections: multipoint, multilinie and multipolygon. OGC provides a standard way to represent geometries through strings, the well-known text or WKT (attribute wkt). A style may determine how the geometry is shown on the map, and teacher can define whether the geographic element is shown on the map or kept hidden (attribute `visible`).

207
Interactions

A map may be associated to an instance of the Interaction class, which represents how the user will interact with it to answer the item.

Some items might include a map, without any interaction required to answer the question (as in the question sketched in Figure 2), and then no interaction is associated to the map.

![Figure 2. Map without interaction required](image)

Four interaction methods are defined:
- **clicking**: clicking on a certain pixel of the map
- **identifying**: clicking on a visible geographic element (marker or geometry) on the map
- **drawing**: drawing a geometry (point, line or polygon)
- **dragging**: moving a geographic element (marker or geometry) to its corresponding place on the map

The method attribute defines which one is used, and its allowed values are respectively click, identify, draw and drag. As the QTI interaction type influences how to answer the question, the interactionType attribute specifies the interaction type defined in the QTI assessment item. Figure 3 shows an example of the same type of interaction method (identifying markers) with two different QTI interaction types: in (a), the learner is prompted to click on the largest Catalan city (a QTI ChoiceInteraction), while in (b), the learner is asked to click on the four Catalan capitals in the order according to their population (a QTI OrderInteraction).

![Figure 3. Interaction method “identifying markers”: (a) ChoiceInteraction, (b) OrderInteraction](image)
An Interaction instance contains one or more validation rules (instances of the ValidationRule class) that represent how the answer is processed to determine whether the answer is right or wrong (or which value is given). The user input is compared to a pattern containing the right answer or one of the possible answers if more than one are considered. The user input depends on the interaction type defined in the Interaction class: a position for click, a geometry for draw, and any geographic element for identify and drag. The pattern is a geographic element, linked by means of the pattern relation with the GeographicElement class. The validation expression (attribute validationExpression) is a string indicating how the user input is compared to the pattern, and it may contain spatial operators and functions to define a rule as precise as needed.

For instance, in Figure 4 (a) the learner is prompted to click on Italy. The interaction method is click, while the validation rule is “PATTERN contains INPUT”, where PATTERN is a constant that refers to the GeographicElement (Italy in this case), INPUT is constant referring to the user input, and contains is a spatial operator. Thus whether the user has clicked inside the Italy polygon is validated.

In Figure 4 (b) the learner is asked to draw the border line between Spain and Portugal. The interaction method is draw but an exact line cannot be expected as answer, and the validation rules contains a margin of error (specified by the teacher) allowed (e.g. 50 km). The buffer spatial function widens the line of the correct response on each side, creating a polygon. The expression of the validation rule is “Buffer (PATTERN,50000) contains INPUT”. Different validations rules (with different validation expressions) could be defined for different margins of error.

Once the expression is validated, it may be true or false, and a corresponding value can be specified (valueTrue and valueFalse attributes), which will be returned to the QTI player. In the Italy example, we can set valueTrue to “Italy” and valueFalse to “Outside Italy”, and define the corresponding QTI assessment item as a choice interaction with two choices: “Italy” and “Outside Italy”.

![Figure 4. Two examples of validation rules (a) Click on a point, (b) Draw a line](image)

**Implementing QTIMaps**

We developed a partial reference implementation of the QTIMaps model, covering the whole lifecycle of edition and runtime. It is focused on Google Maps as it is the web map service best known by teachers and students and contains only the corresponding layers types.

This implementation, called gMapsQTItest (2010), enables us to test the educational benefits and usability of the model, and represents a reference that developers can use to integrate other types of layers covered by the model and other MultimediaInteractions. After presenting the editor and the runtime system, we describe how the tools have been used in two real educational settings.

**The gMapsQTItest Editor**

The editor (see Figure 5) provides a wizard that allows the user to graphically define all the configuration parameters and geographic elements as well as the QTI aspects. The author can choose among a set of pre-defined map
interactions with simple validation rules, such as the click interaction method with a point contained in a polygon as validation rule. After the editing process is finished the corresponding QTI documents (a set of XML files) and additional XML document are generated for each item, describing the corresponding map interaction according to the QTIMaps model. The latter is referenced from the QTI item document by means of an InteractiveMap element. The XML documents can be previewed within the editor and further edited (by experts); and can be published establishing the communication with the runtime system.

Figure 5. gMapsQTItest editor (a) Selection of the interaction (b) Authoring the map interactions (c) Editing the QTI part of the assessment

The gMapsQTItests runtime system

The QTIMaps model has been defined so that available QTI engines and players only require to call some additional middleware when a ‘new’ multimedia element is associated to a QTI item. In the gMapsQTItest runtime system the QTI NewAPIS engine communicates with a middleware that mediates the interpretation of gMapsQTItests by
processing all the elements related to questions dealing with maps. According to the principle adopted in the model, the implementation is based on separating the QTI elements from the issues that are specifically related to maps. NewAPIS processes the QTI XML files associated with the test and the items (Figure 6, step 1). The XML map file is referenced from the itemBody of the question (assessmentItem) through the multimedia interaction element, InteractiveMap in our case (see the red box in 6 - ITEM1.XML). When newAPIS detects a QTImaps element, it calls the middleware (6, step 2) which parses the QTImaps XMLs and generates the code necessary to set the right GoogleMap (6, step 3), the functions that process the user’s actions on the map converting them into QTI responses that can be managed by newAPIS (6, step 4). NewAPIS and the middleware also generate an XHTML code and insert it into a web page that presents the test (6, step 5). This XHTML contains JavaScript functions that call the Google Maps API (6, step 6). The users can interact with the questions and submit the answers which are computed by newAPIS and the middleware (6, steps 7 and 8). Finally the results and feedback are shown to the user (6, step 9).

![Figure 6. Communication between newAPIS and Google Maps using the middleware](image)

It is important to remark that providing support for other mapping solutions, as the WMS standard, would only require middleware modifications, none in newAPIS.

**gMapsQTItest used in learning situations**

To validate the fitness-for-purpose of the QTImaps model from the point of view of educational benefits and usability, it has been used in two authentic learning situations with real users. Both are structured as case studies and analyzed following a mixed evaluation method (Martinez et al., 2003) using several quantitative and qualitative techniques (see Table 2), due to the many contextual factors regarding the students and the teachers.

The questions of the pre-test, the observations and teacher’s comments were used to evaluate the familiarity that students and teachers have with web maps applications, more precisely to answer the question: Did the students demonstrate interest in web map applications? The design of the experiences was targeted to understand the educational benefits of using QTImaps. The questions of the post-test tried to collect data to answer the following research questions: (1) Did the students show a better “attitude to use” gMapsQTItests as compared with a traditional assessments (with blank-maps)? “Attitude to use” refers to the satisfaction/motivation shown by students in the assessment activity. (2) Did the students perceive the gMapsQTItest as more useful than a traditional one? Usefulness refers to the extent to which a student believed that the information provided by the gMapsQTItest would be useful for their learning. Regarding the usability of the QTImaps solution, the data collected in both cases allowed us to answer the following research question: Were the gMapsQTItest interactions easy to use/learn? “Usability” refers to the ease with which participants can use gMapsQTItest.

---

211
Table 2. Data collection techniques

<table>
<thead>
<tr>
<th>Data source</th>
<th>Type of data</th>
<th>Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaires</td>
<td>Quantitative and qualitative participant characteristics, expectations and opinions before interacting with a gMapsQTI test; evaluation of the experience after interacting with the test.</td>
<td>[Pretest] [Posttest] [UPF-test]</td>
</tr>
<tr>
<td>Observations</td>
<td>Notes of observations (opinions done by the participants, behaviour, technical issues, incidents during the activities, etc.) taken by three researchers.</td>
<td>[Observations]</td>
</tr>
<tr>
<td>Automatic Data</td>
<td>Students’ marks obtained in the gMapsQTI test.</td>
<td>[Marks]</td>
</tr>
<tr>
<td>Teacher interviews</td>
<td>Qualitative teacher’s expectations, opinions and concerns before and after the experience.</td>
<td>[TeacherComments]</td>
</tr>
</tbody>
</table>

**Main Case in a Secondary School in Catalonia:** A formal assessment scenario involving a Geography class with 23 students, aged between 14 and 16 years. Several weeks before the exam, the teacher was interviewed in order to know his expectations, doubts and concerns of using an innovative test based on web maps. After presenting him the model and some gMapsQTItests examples in an one-hour session, he designed a test with 8 different questions (see two examples in Figure 7 & 8, and more details of the test, including a video and pictures, can be found at: http://193.145.50.210:8080/QTIMaps/index.html) tailored to the needs of the particular situation. The teacher took into account the students’ marks of the gMapsQTI test to compute the final subject mark (summative assessment). The GMapsQTIEditor was used to create the gMapsQTI test which contained different interactions defined in the QTIMaps model. The exam day, the main features of a gMapsQTI test were explained to the students, who completed the pre-test, the gMapsQTI-tests designed by the teacher and the post-test.

*Figure 7. Question 2. Interaction method: draw. User input: line Validation rule: input contained into polygon*

*Figure 8. Question 6. Interaction method: drag. User input: marker. Validation rule: input contained into polygon*
**Secondary case, an informal open workshop, Science Week at the UPF:** Students from different secondary schools visited an exhibition at the Pompeu Fabra University (UPF) premises during a Science Week. 23 of them completed a gMapsQTItest with 8 questions related to the UPF buildings location using 6 different interaction types. It was an informal spatial thinking assessment scenario. We collected more opinions, expectations, technical issues and concerns of the participants interacting with the gMapsQTItest. As this case cannot be considered a formal e-assessment situation its value is more limited, and the data collected has been used to support or reject the conclusions achieved in the main case.

**Results**

The three main findings and partial results obtained after the evaluation are summarized in Table 3.

First we tried to answer whether students were interested in web map applications. When the teacher was asked about its practices and the familiarity of students with web map applications he said “my students learn geography using their course textbook, blank paper maps and in some cases I show them some maps extracted from Google Earth using the projector. I do not think my students use GoogleMaps at home [TeacherComments]” However, to the teacher’s surprise, 19/23 students had already used Google Maps out of curiosity [Pretest], and indicated that they liked its functionalities because they can explore using the maps. Some uses stated were: “...to learn where different countries in the world are located”; “…to find maps that I don’t find in others sources” and “...to see the countries such as Italy or cities as NewYork, which I like” [Pretest & Observations]. We found out that these students were familiar with web map applications, suggesting that they can use gMapsQTItest quite easily. This is supported by the third finding.

### Table 3. Summary of findings, partial results and supporting data

<table>
<thead>
<tr>
<th>Findings</th>
<th>Partial Results</th>
<th>Support data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The students are interested in web map applications using them to explore geographical information.</td>
<td>The 89% of the students have used web map applications to find answers to their personal interests or to find information that they do not found in the course book..</td>
<td>[Pretest] [Observations] [TeacherComments]</td>
</tr>
<tr>
<td></td>
<td>The students’ familiarity with web maps facilitates the explanations of how to use gMapsQTItest.</td>
<td></td>
</tr>
<tr>
<td>2. The use of gMapsQTItests elicits students’ positive attitudes. The approach seems to also provide significant learning benefits if it is used as formative or summative assessment tool.</td>
<td>Students highlighted that when interacting with web maps: (a) they have more active participation with geographical data, (b) the test is more intuitive, and (a &amp; b) help them to strengthen their geographical memory. Using the gMapsQTItest had a positive effect in their motivation. They like this test because the web maps contain enriched and updated geographical information.</td>
<td>[Posttest] [UPFtest] [Observations]</td>
</tr>
<tr>
<td>3. The students learn quickly how to use the different interactions. The difficulties were on the content of the questions and not related to the interactions themselves.</td>
<td>100% of students in the secondary school passed the test. 5 students had a mark of 4/8 (8 was the maximum score), 6 students had a mark of 5/8, and 12 students had a mark of 6/8. The average mark was 5.30 points. The exam question with less correct answers (only 26.08 %) was 6, which used the “point into polygon interaction using markers”. This interaction type was also used in questions 1 and 3 more successfully answered (95.65% for 3).</td>
<td>[Posttest] [UPFtest] [Observations] [Marks]</td>
</tr>
<tr>
<td></td>
<td>17 out of 46 students (both case studies) selected the “Draw lines and polygons” as the interaction most difficult to use. However, 56.52 % of the students answered correctly the exam question with this interaction and 78.26% of the students got it right in a question of this type of interaction in the informal workshop.</td>
<td></td>
</tr>
</tbody>
</table>
The second finding shows the partial results regarding the educational benefits of the QTIMaps model. The data collected indicated that the gMapsQTItest elicited a positive students’ attitude, and that gMapsQTItests could be useful to learn and assess geographical skills. The teacher was satisfied with the designed gMapsQTItest, and especially glad about the accuracy of the answer computation of the feedback [TeacherComments]. Learners were somehow scared before the test “They seem very excited” [Observations], since their results would be part of the final mark, but not because of using technologies or an unfamiliar scenario. Despite it, there were positive comments after the gMapsQTItest, “... it is funny and useful to learn!”, “… I like this experience because normally we take exams on paper” [Posttest]. Explicit advantages for learning geographical stated were: “…I like it, because it doesn’t have limits and you have to think much more...It is more complete (than a traditional test with blank maps)”, “…it has been an amusing activity and it helps you to remember better”, “… I prefer these maps because I can interact with the computer” [Posttest]. This seems to indicate that interactions demand more active students’ participation helping them to remember better the concepts, in line with psychological studies demonstrating that visual images (as web maps) have a beneficial impact in the memory ability of students (Wager, 2005). Moreover, the informal workshop data support these findings with participants’ comments as: “I think that it is an interesting and enjoyable method to study geography (I hated this subject...)”, “...it is a different way to do an exam, because you enjoy doing these tasks and this helps to remember better your actions” [UPFtest].

The third results are about usability; in summary, users can learn and use gMapsQTItests quickly and effectively. Students said: “… I don’t have difficulties to answer any question”, “… all the functionalities (such as zoom, pan or drag markers) are very easy!” [Posttest]. Exam question 6 (see Figure 8) was the one students failed most (17 of 23), but they got good results in questions 1 and 3, which required the same interaction type (dragging a set of markers to specific positions). This indicates that interaction and content difficulties are dissociated, which agrees with the teacher comment that question 6 was more difficult [TeacherComments]. This is supported too by the fact that the most difficult interaction type selected by the participants, “Draw lines and polygons”, namely by 17 of the 46, (see Figure 7) did not have an impact in questions using it having worse answers. Furthermore all the learners passed the exam (the average mark was 5.30 out of 8) [Marks]. These results, combined with the familiarity with web maps revealed in the pre-tests, showed that students understand all the functionalities in gMapsQTItest and can criticize problematic interactions, their opinions being very relevant. Overall, we can say that learners interacted with QTIMaps questions implemented in the test without difficulties of use.

Other learning scenarios regarding the assessment of web map skills

The two previous scenarios cover Rhind’s first and second types of questions for geographical skills (see Table 1). In this section we present other assessment scenarios that validate that the QTIMaps model is able to cover the rest of Rhind’s types, and other disciplines. These scenarios have been proposed in a recent on-line course for teachers on integrating on-line geographical resources of the Balearic Islands in the K-12 curriculum (Cavada & Navarrete, 2010).

Reasoning about changes in the territory of tourist areas

In this scenario the Balearic Islands SDI (IDIEIB, http://www.ideib.cat) is used to analyse how the territory has changed from the situation before the tourist boom of the 1960’s. The proposed activity consists in identifying which urban areas in the coast of a municipality of Minorca did not exist 50 years ago. Three WMS layers are used: the 2008 aerial photography (the last available), the one made in 1956 (digitized recently), and the municipalities’ borders. The learners have to draw the polygons of the areas that have appeared, which will be automatically compared to the correct ones. The map interaction is modelled as a draw method, where the user answer is a set of polygons and each of them is checked deciding whether it is contained inside the right one (applying a buffer to allow a margin of error).

This is an example of a formative assessment activity, where the results of the test will be the basis for debating how tourism has transformed the environment. In this scenario the third Rhind type of question is used.
Reflecting about the Balearic fauna and flora space

In this scenario, the learner is asked to find the areas where certain fauna and flora species coexist in the Balearic Islands. This helps them to identify the typical ecosystems of a Mediterranean island.

This is an example of finding spatial patterns (4th Rhind’s question) within a formative assessment activity, where several WMS layers about biodiversity from the IDEIB are used. The type of interaction is similar to the previous one, where the learner has to draw polygons where the species coexist. In geographical terms, the species coverage overlap, and these correct solutions can be automatically computed and compared to the answers provided.

Modelling a new railway line

The learner is asked to propose the course for a new railway line in Majorca connecting the Alcúdia Bay to the current network. The learner is given a simplified model including some of the constraints that the line should satisfy. For instance, it should not go through protected natural areas and it should run close to existing roads. S/he is asked to draw the new line on the map, which will be automatically compared to the two main real alternatives. Draw is the interaction method with a line as user input. This scenario is an example of modelling, 5th Rhind’s question, and the assessment can be carried out either as formative or summative. Different WMS layers from the IDEIB are used in this scenario, such as railway and road networks, protected natural areas and recent aerial photography.

Conclusions

The integration of interactive maps in learning technology solutions facilitates new ways to assess relevant geographical skills, such as spatial thinking. The QTIMaps model we introduced extends the QTI standard with new interaction types for answering items that exploit the interactions supported by web maps, so that new types of automatically corrected questions can be created. We showed that the model is applicable and supports a variety of assessment scenarios, both old and new. An implementation of the model, using the widely popular Google Maps as web map service and only requiring some specific middleware beyond a QTI engine has been presented. It has been used in two real and different educational situations showing educational benefits going beyond assessment – such as students’ motivation increase and memory reinforcement, related to students having to interact actively with web maps to answer the questions – and the usability of the editor and the player. Three additional assessment activities framed in different educational scenarios were described, and showed the potential of the use of the model, as well as the use in the experiences. They cover geography and other fields where georeferencing is relevant.

The QTIMaps model contributes to the state of the art in different lines. The results of the two experiences indicate that the new tools provide a better user experience. Web maps offer navigation possibilities, such as pan and zoom, which allow the learner to contextualize a question in the world (although this can be limited by the teacher) and encourage exploration. Web map viewers as Google Maps provide a familiar user-friendly interface and its use removes barriers perceived by the students between classroom and outside world. The use of OGC compliant web maps gives access to an enormous source of layers with accurate information of different geographically related topics published through SDIs. Moreover, the information is continuously updated by the organizations that produce it, and consequently, once the teacher selects the map for the assessment, the map will always remain up-to-date. The combination of spatial relations and operators allows the teacher to define complex interactions that cannot be done through the existing QTI graphical interactions. Let us remark that the QTI graphical interactions – such as, for instance, SelectPointInteraction - use pixel coordinates and if the image in the question is changed or viewed in a different way, the interaction has to be re-defined again for the pixel coordinates of the new image. In our approach, the coordinates of the geographic elements do not change when the map appearance changes. This lack of flexibility and of semantics, provided by the information layers of maps, are important reasons to introduce and use the geographic interactions and not keep the QTI graphic interactions. While obtaining and publishing an appropriate map in a QTI graphical interaction can be a hard task, the use of web maps makes it much simpler: since the information is already available on the web, the teacher just needs to select the layers and the area of interest. In this paper, we have shown that teachers have been able to design assessment scenarios based on the model after short sessions where the model, the tooling and some examples were presented.
Our future plans include carrying out a more extended evaluation of the use of the editor with teachers in order to measure the effort required to learn and apply the model. We also intend to combine the QTIMaps model with other learning standards such as IMS LD or SCORM. This combination will allow teachers to include the assessment of geographical skills in their learning designs using educational technology standards in the whole cycle.

The IMS Tools Interoperability working group (IMS, 2007), is devoted to enhance the interoperability of tools within Learning Management Systems, and our approach can be seen as contributing to the cases studied by the group. First, our implementation based on a middleware communicating web maps services and a QTI engine under a strict separation principle can be easily generalized as facilitating interoperability. Second, the model itself could be extended to include other multimedia interactions, beyond web maps. Our future work considers extending it to those types of interactions related to time-based graphical representations (e.g. the multimedia service offered by the TimeLine project, SIMILE, 2003).

Acknowledgements

This work has been partially funded by the Spanish Learn 3 project (TIN2008-05163/TSI). The authors would also like to thank the support of the GTI members Javier Melero and Mar Pérez-Sanagustín, and the DUC de Montblanc High School.

References


Learning paths can be roughly defined as sets of one or more learning activities leading to a particular learning goal. Introduction

Keywords


Evaluation of the Learning Path Specification

José Janssen, Adriana J. Berlanga and Rob Koper
CELSTEC Open University of the Netherlands, Heerlen, The Netherlands // jja@ou.nl // abl@ou.nl // rkp@ou.nl

ABSTRACT

Flexible lifelong learning requires that learners can compare and select learning paths that best meet individual needs, not just in terms of learning goals, but also in terms of planning, costs etc. To this end a learning path specification was developed, which describes both the contents and the structure of any learning path, be it formal, non-formal, informal, or indeed a combination of these. This paper briefly explains the learning path specification and presents a framework for the evaluation of the specification based on theories of model quality. A study of learner choice processes (n=15) was carried out to investigate the specification’s semantic and pragmatic quality (clarity, completeness and parsimony) with respect to the selection of a learning path. Results indicate that the specification does not contain any redundant information. Instead, the study has led to improvement of the specification’s (feasible) completeness by further refinement of scheduling information.

Keywords

Formal learning, Informal learning, Learning path specification, Learner choice

Introduction

Learning paths can be roughly defined as sets of one or more learning activities leading to a particular learning goal. Learning paths can vary from a relatively small activity like reading a book or taking a course to following an entire programme or curriculum. Learning paths may vary also regarding the level of formality. In line with the Commission of the European Communities we distinguish formal, non-formal and informal learning (CEC, 2000).

Whereas formal learning occurs in education and training institutions and leads to recognised diplomas and qualifications, informal learning is described as “a natural accompaniment to everyday life” which is not necessarily intentional learning (CEC, 2000, p. 8). Non-formal learning, finally, is learning that takes place alongside the mainstream systems of education and training, for instance at the workplace or in arts or sports, which does not necessarily lead to formalised certificates.

Lifelong learners’ learning paths consist of a mixture of formal, informal and non-formal learning (Colardyn & Bjornavold, 2004; Colley, Hodkinson, & Malcolm, 2003; Livingstone, 1999). In order to support lifelong learners in comparing and selecting suitable learning paths, a uniform way to describe learning activities and learning paths has been developed, which covers these different ways in which people learn (Janssen, Berlanga, Vogten, & Koper, 2008).

The specification is envisaged to support several processes. Firstly, it is meant to be used by educational providers to describe formal and non-formal educational courses and programmes in order to make them available through specific search engines, thus enabling comparison across providers. We assume that educational providers will want to describe learning paths in a uniform, formalised way, because the benefits of transparency and opportunities for automated learner support outweigh the costs. Costs can be relatively low since educational providers already have to describe their offerings; it will merely be a matter of organising this information in a way that enables storage and update in one place and subsequent use in different contexts: printed catalogues, websites, and search engines.

A second process the learning path specification is meant to support was initially defined as follows: lifelong learners use the specification to describe their informal learning paths to make them available as an example to other learners with similar learning goals. However, a pilot-study revealed that it requires considerable efforts and skills on the part of the learner to identify activities that did or did not after all contribute to achieving those outcomes. To distil a learning path from one’s own informal learning experiences and describe it in a way that is useful for others, is not an easy task (cf. Skule, 2004). Though we still maintain that the specification can be used to describe all kinds of learning (a point we later further elaborate), we believe that in the case of informal learning it is not likely going to happen on a large scale by lifelong learners themselves, because it requires learning design skills. It is not unreasonable though to expect employers and employment agencies to be willing to invest in these descriptions of informal learning paths as they can offer tried and tested alternatives to more costly formal and non-formal learning.
paths. Research indicates that people spend an average of 6 hours a week on employment related informal learning (Livingstone, 1999) and description of these informal learning paths is likely to enhance efficiency when they can offer guidance to learners rather than have them find things out through trial and error. In any case, the second process the learning path specification is meant to support eventually is defined as: description of informal learning paths in order to make these learning paths available for other learners with similar learning goals.

Finally, a third process the learning path specification is envisaged to support is selection of suitable learning paths. To this end the specification identifies main characteristics to be used in comparing and selecting a learning path (e.g. learning objectives, prerequisites, study load, costs, et cetera). Lifelong learners must be offered means to efficiently choose the learning path that best fits their needs. Taking a decision support perspective, we distinguish two stages in this process: screening and choice (Beach, 1997; Rundle-Thiele, Shao, & Lye, 2005). Screening involves selecting a number of options one wants to take into consideration, i.e. narrowing down the number of choice options to a number that can be “managed”. Research shows that choice overload may occur due to the number of available options, as well as to the number of attributes related to these options (Fasolo, McClelland, & Todd, 2007; Malhotra, 1982). In other words: having to choose from a large number of learning paths is one thing, having to compare even a limited number of learning paths might lead to choice overload when a large number of attributes are related to these options. But even apart from these considerations regarding choice overload, lifelong learners will rather invest the scarce resources of time and attention in developing competences than in comparing all kinds of ways to do so. What is needed then is some tool for the learner to select a limited set of learning paths to take into account in the choice process.

There are quite a number of criteria that could be relevant to finding the most suitable learning path but not all criteria might be equally relevant to all learning paths or to all learners for that matter. The study of Fasolo, McClelland and Todd (2007, p. 23) shows that “it is possible for consumers to make good choices based on one or two attributes, when attributes are positively related or consumers care unequally about attributes and choose on the basis of the most important ones”. To the extent that learners do not equally care about the learning path attributes included in the learning path specification progressive disclosure of functionality could contribute to help the learner focus on those criteria that are most relevant for her (Turbek, 2008). Progressive disclosure is a strategy for managing information complexity in which only necessary or requested information is displayed at any given time (Lidwell, Holden, & Butler, 2003).

Requirements for the specification have been derived from a review of literature on curriculum design and lifelong learning as well as observations of current practices to support learner choice (Janssen et al., 2008). This paper describes a study directed towards evaluation of the conceptual model of the learning path specification. It provides an outline of the specification and explains how the specification supports description and selection of learning paths. Subsequently a framework for the evaluation of model quality is presented, guiding the specific research questions. Finally the paper describes method and results of the evaluation.

**Learning path conceptual model and specification**

According to Moody (2005) conceptual modelling is a process of formally documenting a domain (a system or a problem) in order to enhance communication and understanding. He further points out that conceptual modelling may be used to describe requirements at different levels: functional and non-functional requirements at the level of an application, and information requirements at the level of an organisation or even an industry.

A formal specification can be considered a conceptual model as is illustrated by the following definition: “a formal specification is the expression, in some formal language and at some level of abstraction, of a collection of properties some system should satisfy” (Van Lamsweerde, 2000).

The learning path specification identifies information requirements for lifelong learners: generic elements of a learning path which are essential to selecting, planning and executing a learning path, such as learning goals, learning actions, delivery mode, etc. It describes fixed as well as optional elements; both contents and structure.

Like any other path a learning path has a finish and a start (i.e. learning goals and prerequisites). In order to get to the finish one or more learning actions have to be completed. Learning goals and prerequisites can be specified both at the level of the learning path and its constituent actions. They are preferably defined in terms of standardized
competences so as to facilitate automated identification of parts of a learning path a learner may skip when these competences have already been attained through prior learning (Kickmeier-Rust, Albert, & Steiner, 2006). Learning actions can be grouped into clusters, for instance when they compose a set a learner can choose from (selection), or because they have to be studied in a particular order (sequence). Table 1 describes the information about the learning path and its constituent actions considered relevant in identifying and selecting suitable learning paths, and therefore included in the specification. (For a more detailed account of the attributes and metadata associated with each of the elements see Janssen, Hermans, Berlanga, & Koper (2008)).

**Table 1. Learning Path metadata**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Name of the program, course, workshop etc.</td>
</tr>
<tr>
<td>Description</td>
<td>Brief description of the program, course, workshop, etc.</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>Competences which are expected to have been acquired beforehand</td>
</tr>
<tr>
<td>Start conditions</td>
<td>Other conditions that must be met in order to start: e.g., a minimum number of participants, a special diploma, access to a computer, microscope…</td>
</tr>
<tr>
<td>Language</td>
<td>Languages used in the learning path</td>
</tr>
<tr>
<td>Diploma/certificate</td>
<td>Indicates whether completion of the learning path results in an officially recognized diploma or certificate</td>
</tr>
<tr>
<td>Time investment</td>
<td>Total number of hours it takes to complete</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Indicates whether the learning path involves self-study, face-to-face meetings or a mixture of these.</td>
</tr>
<tr>
<td>Guidance</td>
<td>Description of available guidance</td>
</tr>
<tr>
<td>Assessment</td>
<td>Description of assessments associated with the learning path</td>
</tr>
<tr>
<td>Start date/end date</td>
<td>Start/end date</td>
</tr>
<tr>
<td>Costs</td>
<td>Total costs for enrolment, materials, etc.</td>
</tr>
<tr>
<td>Number of contact hours</td>
<td>Indicates number of hours of required (virtual) presence.</td>
</tr>
<tr>
<td>Location</td>
<td>Indicates where meetings take place</td>
</tr>
<tr>
<td>Completion</td>
<td>Indicates how and by whom it is decided whether the learning path goals have been achieved.</td>
</tr>
<tr>
<td>Provider</td>
<td>Provider of the learning path</td>
</tr>
<tr>
<td>Further information</td>
<td>Link to a website for further details.</td>
</tr>
</tbody>
</table>

**Formal, non-formal and informal learning paths**

The distinction between formal, non-formal and informal learning is not as clear-cut as the definitions provided in the introduction suggest. Schugurensky (2000) stresses the fact that informal learning can also take place inside formal and non-formal educational institutions: within these institutions some learning occurs independently of the intended goals of the curriculum. Using two categories (intentionality and consciousness) he goes on to identify three forms of informal learning: self-directed learning (intentional + conscious), incidental learning (unintentional + conscious) and socialization (unintentional + unconscious). The learning path specification is merely meant to enable description of informal learning with the aim to suggest informal ways to develop competences, drawing from other learners’ personal informal learning experiences. This means the learning path specification is only meant to cover conscious informal learning. As to the intentionality of learning it is often stated that workplace learning and other informal learning have no formal curriculum or prescribed outcomes (Hager, 1998). Regarding unintentional conscious learning we maintain that this type of learning can be described in hindsight as a learning path, describing the previously unintentional learning outcomes as learning objectives, to present as an option to other learners interested in achieving these learning objectives.

Concerning the distinction between informal and non-formal learning, a major review of literature suggests there is no clear agreement: the terms are used interchangeably (Colley et al., 2003). Nor does it appear possible to distinguish formal learning from other learning in ways that have broad applicability or agreement. The authors conclude it is more sensible to consider attributes of informality and formality present in all learning situations. These attributes concern four aspects of learning:
• Process: informality and formality attributes relating to the learning process relate to questions like who’s in control of the process (teacher controlled versus student led), whether and what kind of assessment is involved (formative or summative).

• Location/setting: where does the learning take place (e.g. in an educational institution, at the workplace, etc.) and does it involve certification?

• Purposes: is learning intended or does it happen unintentionally; are learning outcomes determined by the learner or designed to meet needs which are externally determined?

• Content: does the learning focus on acquisition of established knowledge or development of knowledge from experience?

Attributes relating to the process aspect of learning included in the specification are the metadata elements “guidance” and “assessment”. The location/setting aspect is covered by the metadata “recognition”, “delivery mode”, and “location”. Regarding the purpose aspect we conclude that the learning path specification only covers intentional learning: a learning path is directed towards learning goals. This does not mean that the learning path specification cannot be used to describe unintentional learning as well, but this would always be in hindsight: learning which has occurred unintentionally can be retrospectively described to serve as an example to other learners who can then embark on the same path purposefully. Attributes of formality and informality relating to the content aspect of learning can be described through the metadata element “description” of the learning path as well as of its constituent actions. Whether the learning goals of a learning path are achieved through “formal knowledge acquisition” or through “learning by doing” will be of interest to the learner, but whether it requires a separate metadata element remains to be seen.

Model evaluation: a framework

Seeking alignment with the ISO 9000 definition of quality Moody (2005) defines conceptual model quality as “The totality of features and characteristics of a conceptual model that bear on its ability to satisfy stated or implied needs” (p. 252). Based on a review of research in the field of conceptual model quality Moody concludes that there are no generally accepted guidelines for evaluating the quality of conceptual models. Nor do experts agree as to what makes a conceptual model a “good” model. One of the explanations given for this lack of consensus is that a conceptual model exists as a construction of the mind, and therefore quality of a conceptual model cannot be as easily assessed as the quality of a concrete product: “While the finished product (the software system) can be evaluated against the specification, a conceptual model can only be evaluated against people’s (tacit) needs, desires and expectations. Thus the evaluation of conceptual models is by nature a social rather than a technical process, which is inherently subjective and difficult to formalise” (Moody, 2005, p. 245).

The learning path specification is a case in point: rather than a “finished product” it is a model to describe learning paths which can be used to develop tools to support lifelong learners in finding and navigating suitable learning paths. This implies a number of stakeholders:

- lifelong learners
- learning path designers
- providers
- software developers.

Someone interested in finding suitable learning paths is likely to focus on different aspects of the learning path specification than someone interested in designing learning paths or in developing tools to support these processes. Consequently, evaluation of the specification requires input from these different perspectives.

Addressing the lack of consensus in the field Moody (2005) proposes the ISO/IEC9126 software quality model as a template to structure conceptual model quality frameworks. This template identifies the following important features:

- hierarchical structure of quality characteristics (characteristics, sub-characteristics and metrics)
- familiar labels
- concise definitions
- measurement (characteristics are operationally defined)
- evaluation procedures (who should be involved how and when).
Concerning the hierarchical structure of quality characteristics, we will draw on a distinction which, despite the observed overall lack of consensus, several researchers in the field adhere to (albeit not exclusively): syntactic quality, semantic quality and pragmatic quality (Krogstie, 1998; Leung & Bolloju, 2005; Lindland, Sindre, & Solvberg, 1994; Moody, Sindre, Brasethvik, & Solvberg, 2002; Recker, 2006; Teeuw & Berg, 1997).

The framework for the evaluation of the learning path specification we developed is presented in Table 2.

Syntactic quality involves the extent to which the conceptual model adheres to the syntax rules of the language it is modelled in. In the case of the learning path conceptual model evaluating the question would be whether UML has been properly used (i.e., in accordance with UML syntax rules) to express what was meant to be expressed.

Semantic quality refers to the extent to which the model accurately represents the essential features of the phenomenon under study. Some of the differences in defining model quality revolve around the interpretation of what constitutes an accurate representation. Interpretations of accuracy vary, depending on whether or not the phenomenon under study is considered an “objective reality” (ontology), and whether or not it is possible to objectively know this reality (epistemology) (Recker, 2005). Regarding semantic quality several authors mention specific criteria like completeness, validity, clarity, consistency, etc. (Krogstie, 1998; Leung & Bolloju, 2005; Recker, 2005; Teeuw & Berg, 1997). However usage of these criteria is not consistent. Moody et al. (2002) for instance use the term validity to indicate a number of criteria (completeness, parsimony, and independence) which others use to define semantic quality. Interestingly, Krogstie (1998) introduces the notion of feasibility. Whereas completeness means that the model contains all the statements which are correct and relevant to the domain, feasible completeness means that there are no statements in the domain, and not in the model, which would be cost-efficient to include. Besides, this author distinguishes between semantic quality and perceived semantic quality. He argues that the primary goal for semantic quality is for the model to correspond with the domain. However, this correspondence can not be checked directly since:

“To build a model, one has to go through the participant’s knowledge regarding the domain, and to check the model, one has to compare with the participant’s interpretation of the externalized model. Hence, what we observe at quality control is not the actual semantic quality of the model, but a perceived semantic quality, based on comparisons of the two imperfect interpretations” (Krogstie, 1998, p. 87).

Pragmatic quality finally refers to the question whether/how easily the model is comprehended by the stakeholders in view of its purpose. The purposes of conceptual models can vary widely: enhance communication, document the current state of knowledge, guide system development, exploration, prediction, decision support (Beck, 2002; Moody, 2005). Pragmatic quality can be further split into technical pragmatic quality and social pragmatic quality (Nelson, Poels, Genero, & Piattini, 2005), indicating whether the model is easily interpreted by tools and human users respectively.

Research questions

Syntactic quality has been evaluated mainly through peer review and expert consultation. So far the model mainly has been used for communication purposes. Eventually the UML model will be transformed to an XML schema which requires greater refinement and detail, and further evaluation of syntactic quality. This evaluation will be reported about in a separate publication.

Semantic quality has been evaluated through collaboration with software developers and processes of peer review. However the elements and characteristics identified by the model have been derived from a review of literature and current practice, but are these really the elements and characteristics lifelong learners want to be informed about? Are these the elements and characteristics they take into account when considering different options?

<table>
<thead>
<tr>
<th>Quality dimensions</th>
<th>Description</th>
<th>Sub-characteristics</th>
<th>Description</th>
<th>Evaluation methods</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic</td>
<td>Does the</td>
<td>Proper notation of</td>
<td></td>
<td>- submit model to</td>
<td>- number and</td>
</tr>
<tr>
<td>Quality</td>
<td>Model correctly express what is meant to be expressed in accordance with UML syntax rules?</td>
<td>Association, aggregation, generalization, multiplicity etc.</td>
<td>Peer/expert review - validity checks through software</td>
<td>Type of errors, ambiguities etc.</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Semantic quality</td>
<td>Does the model represent essential features of the phenomenon under study?</td>
<td>Adequate [1], orthogonal/independent [3, 5], valid [2, 4]</td>
<td>The model adequately reflects the domain, i.e., independent aspects are captured by different concepts and relations are adequately represented.</td>
<td>- Explain the model to lifelong learners and learning path providers to see whether they find it adequate on points relevant to them - Analyse lifelong learners’ learning path choice processes to establish learning path characteristics essential in this process - Map existing learning paths on model</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complete [1, 3, 4, 5], nothing missing what is expected [2]</td>
<td>The model describes all essential features.</td>
<td>- Number and type of issues open to debate - Number of changes made to the model - Number and type of frictions in mapping learning paths</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimal [1], parsimonious [3, 5], nothing unexpected presented [2]</td>
<td>The model does not contain irrelevant aspects and relations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pragmatic quality</td>
<td>Is the model easy to understand?</td>
<td>Unambiguous [1, 3]</td>
<td>Concepts and relations have a clear single meaning</td>
<td>- Perceived ease of use - Perceived usefulness - Intention to use</td>
<td></td>
</tr>
<tr>
<td>Social &amp; Technical</td>
<td></td>
<td>Internally consistent [1, 3]</td>
<td>- Establish whether the specification is adequate to develop tools. - Establish whether tools developed are considered useful</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>General</td>
<td>Concepts should be as independent as possible from any specific application (domain)</td>
<td>- Map informal and non-formal learning paths from different domains</td>
<td></td>
</tr>
</tbody>
</table>

Evaluation of pragmatic quality will focus on software developers and tools. However in our view it makes sense only to evaluate pragmatic quality after semantic quality has been sufficiently tested, because poor semantic quality will inevitably result in poor pragmatic quality. Still some aspects of pragmatic quality will be included in the present study as well, involving the question whether the learning path characteristics included in the specification are clear and easy to understand.

More particularly, the focus of the present study is on the following quality aspects relating to the purpose of enabling comparison and selection of learning paths:

1. Is the information provided by the model clear? (pragmatic quality)
2. Is the specification complete: does the model contain all essential information lifelong learners desire/need to select suitable learning paths? (semantic quality)
3. Is the specification minimal: does the model contain information which is not considered relevant by lifelong learners? (semantic quality)

Method

Above research questions were addressed through a case study examining lifelong learners’ decision making processes (Flyvbjerg, 2006; Yin, 2003). Data on decision making processes were gathered through semi-structured interviews with learners (n=15) who recently chose a learning path, having considered at least two different options. Participants for the study were recruited asking colleagues and acquaintances to propose candidates from their network of family and friends.

Typically sampling for multiple-case studies is guided by the research questions and conceptual framework. Our main sampling strategy was maximum variation of cases (Flyvbjerg, 2006; Miles & Huberman, 1994), meaning that we sought to include a broad variety of learning paths regarding domains of personal/professional development, and level of formality. Besides we aimed to have a broad variety of respondents regarding age, gender, employment status, and prior education. The number of cases to include was not pre-determined, but including over about 15 cases is acknowledged to make it harder to keep an overview without losing sight of necessary details (Miles & Huberman, 1994). Though essentially each case has unique properties and is therefore interesting in its own, in hindsight it appears that the last four interviews did not provide any new information regarding the characteristics taken into account in the decision making so that in this respect a point of saturation (Miles & Huberman, 1994) seems to have been reached. The risk of retrospective distortion due to inaccurate recall was reduced by requiring that the decision making process had come to a conclusion no longer than three months ago, and by using a technique of aided recall during the interviews (Coughlin, 1990).

Interview protocol

The interview protocol included four steps. First participants were asked to tell a bit more about their motives to learn. The second step focused on spontaneous recall: participants were asked to describe their search for ways to achieve these learning goals and how they “weighed” these different options, i.e., on which characteristics they compared them to arrive at a final choice. Any characteristics mentioned during the interview which were not part of the learning path specification were noted down by the interviewer. The third step involved aided or prompted recall: participants were invited to go through a set of cards, each card containing a label and description of a characteristic included in the specification as shown in Table 1, complemented with two additional cards for learning outcomes (knowledge and skills to be developed) and learning actions (things you have to do: study, investigate, write, present, etc.).

For each of the cards participants were asked to indicate whether the described characteristic was clear to them and whether it had played a role in the recent choice of a learning path. The fourth step required of participants that they shift from the most recent decision making process to deciding on a learning path more generally, and to consider whether in general they would want to take this information into account.
Cases

Figure 1 presents the learning paths included in the study classifying them along two dimensions: relation to career and “urgency”, i.e., the question whether the learning path is considered a “must have”. This second dimension emerged as a relevant distinction during the interviews: whether or not the learning path is conditional, i.e. whether it enables the learner to do things which will otherwise remain beyond her reach (e.g. apply for another job). Though at face value one might expect conditional learning paths to exist mainly in the realm of professional development, there are several counter examples, such as learning to swim or drive a car. In the case of the career related learning paths the conditional learning paths were “must haves” either with respect to adequate job performance, or to a job or career switch. Other career related learning paths were merely meant to “look good on the CV”, without an immediate urge to find another job.

Results

The number of learning paths compared in depth in the decision making processes varied between 2 and 8, with an average of 4. In twelve cases Internet was used to search for suitable learning paths. Two cases involved a restricted choice between two options offered by the employer or educational institution. In a number of cases the process of screening had started about a year before. The distinction between screening and choice is not as clear-cut in practice as in theory: rather there exists a grey area of learning paths which are considered more closely but still get dropped long before the final choice is made. A clear distinction between screening and choice can be made only in those cases where one or two criteria stand out as initial selection criteria as was, for example, the case with the choice of a driving school, where a first selection (screening) took place on the base of reputation (pass/fail rates) and location.

An interesting general observation regarding the in-depth comparison leading up to the final choice is that in the case of the informal learning paths the choice process entailed some probing of different options. Of course this was possible because these options were freely available and did not require any formal subscription or enrolment. However they were nevertheless considered as clearly distinctive options: though there was a period of “trial” eventually a choice for a particular option was made, rather than for a mix.
Spontaneous recall

Figure 2 shows - in descending order - to what extent learning path characteristics played a role in the decision making process according to the spontaneous recall of participants. The characteristics “title” and “description” have been left out, as they are obvious. Characteristics which were mentioned during the interview and which were not included in the learning path specification are marked by (+).

Some caution is required regarding the interpretation of these results. All participants were more or less aware of the learning outcomes of the learning paths under consideration but they did not always play a role in the comparison, simply because the learning paths were more or less identical in this respect, or because the learning outcomes were less important than acquiring the associated diploma or certificate. Similarly, language was not mentioned as a criterion in the decision making process simply because all learning paths considered were in Dutch. In these cases the characteristic has played an (implicit) role in the process of screening.

Contact time, experience/advice, quality, and teacher were mentioned in addition to the characteristics included in the specification and merit closer inspection.

Experience/advice: six participants remarked they had been keen to acquire information on other peoples’ experiences concerning the options they were considering. Preferably people they were acquainted with so that their judgement could be appraised, but otherwise in the form of Internet forums.

Teacher: three respondents compared information on the teacher involved, placing different accents: two were merely interested in teaching experience (number of years) and the third considered it very important that the teacher had practical work experience in the subject area (Law).

Contact time: contact time involves the question at what time of the week face-to-face meetings take place. Scheduling information is multi-faceted as is already expressed by a number of characteristics included in the specification: start/end date, delivery mode (contact: yes/no), and contact hours (amount of contact). Now additional information is called for regarding the time of the week contact takes place. The indication “part-time/full-time”
which is sometimes used was not included in the specification because it is too general to be informative. This is confirmed by the specifications from participants in this study: “not on Wednesdays”, “only evenings or weekends, depending on how far I have to travel”, etc. What is required is a categorisation that is specific enough to be informative, yet general enough to be practical.

Quality: five respondents said they had taken into account the quality of learning paths. When asked how they had established quality, a variety of aspects was mentioned: pass/fail rates, “does the website look professional”, and quality of learning materials (e.g., up-to-date content).

Aided recall

Despite the brief explanation offered on the cards the characteristics were not always clear and unambiguous. However, this seemed somewhat intrinsic to the domain as several characteristics included in the specification are closely related, nuances tended to get lost, for instance, regarding the concepts “assessment”, “completion”, and “recognition”. Assessment describes the types of assessment(s) included in the learning path, and completion indicates whether there is a formal end to the learning path (set by an assessment or time limit for instance) or whether it is up to the learner to decide whether the learning goals have been reached. Though both concepts are clearly related to recognition, they are not identical: recognition is independent of types of assessment and does not necessarily mean deadlines.

Also, though characteristics themselves may be clear and unambiguous, the role of the characteristic in comparing and selecting learning paths may not be unambiguous. Indeed plain and simple characteristics like costs and time investment could lead respondents to ponder: of course, generally speaking, you would want to reduce costs as much as possible, but then again “quality comes with a price”.

Figure 3 compares the results for spontaneous recall (s) with the results based on aided recall (a).

![Figure 3. Spontaneous (s) vs. aided (a) recall](image-url)
Clearly none of the characteristics included in the learning path specification can be considered superfluous. Apparently quite a number of characteristics are prone to be overlooked in spontaneous recall. In fact, only the results for outcomes and location appear remarkably stable. Figure 3 serves to illustrate how certain characteristics are more often taken into account in the process of selecting a learning path than reports based on spontaneous recall would suggest. Some of these characteristics were taken into account implicitly, without the learner being consciously aware of it (e.g., delivery mode). In other cases the characteristics had been consciously considered, and subsequently forgotten as they had not constituted an issue: “Yes, I do recall looking at guidance information, but it was ok…”.

Several respondents commented that they had not seen any information regarding certain characteristics (e.g. assessment, actions, prior knowledge, and guidance). Thus results may to some extent reflect the availability of information. Figure 4 confirms that in general a majority of the learners want to be informed on each characteristic when deciding upon a learning path.

Discussion

This section discusses the question whether the characteristics mentioned by respondents in addition to those included in the specification, should be added. Several participants started out selecting learning paths based on provider names with solid reputations. However, even in these cases additional information was sought on learner experiences with these learning paths. However, information on learner experiences can not be included in the specification, because the description of learning paths is made by the provider and the information on experiences is only of value when it is completely independent. Alternative solutions might be found in adding annotations or ratings provided by users, or in providing recommendations through collaborative filtering, e.g., “Your profile most closely matches the profile of learners choosing learning path X” (Drachsler, Hummel, & Koper, 2008). However, participants expressed a preference to hear about experiences from people they know so as to be able to appraise their judgement. Further research is needed to establish whether the proposed solutions are viable alternatives.
In three cases information was sought on the teacher (number of years in teaching or practical professional experience in the subject area). The question is whether this information should be provided through one or even two separate characteristics in the specification, or whether this is typical information a learner should be able to find through the link provided via “further information”. Though teacher information can be decisive, it will hardly play a role at the stage of screening but rather towards the end of the process in the comparison of a limited set of options. This is not the case for the information regarding contact time, i.e., the scheduling of meetings associated with a learning path: this information will help to distinguish suitable learning paths at the very start of the decision making process. Including this element in the specification is therefore likely to contribute considerably to efficiency. So bearing in mind the notion of feasible completeness the element “contact time” will be added to the specification. Seeking a balance between the level of detail some participants described and considerations of what is practical, two dimensions will be distinguished: weekdays/weekend and daytime/evenings.

Finally, the aspect of “quality” was mentioned, referring to a variety of indicators: pass/fail rates, a probe of learning materials (up-to-date), or impressions of professionalism. This type of information can not be grasped simply by adding another learning path characteristic, but has to be sought in addition, through independent sources.

**Conclusion**

We investigated 15 choice processes involving a broad variety of learning paths, with the aim to evaluate semantic and pragmatic quality of the learning path specification: are characteristics included in the specification to support comparison and selection of a learning path clear, sufficient, and without redundancies?

Regarding clarity our study showed that related characteristics (e.g., delivery mode and contact hours) sometimes got mixed up. However, this can be solved by presenting them in combination and with possible values.

None of the characteristics included in the specification appeared redundant. Instead, several characteristics were mentioned in addition. However, upon closer inspection, only contact time appears an adequate improvement of the specification.

Following this investigation and adaptations made on the base of these results, a tool is being developed to describe learning paths in line with the specification. Subsequent tools can then be developed which use these descriptions to facilitate selection of suitable learning paths.

Though several participants hinted at information overload regarding the number of learning paths, one respondent specifically hinted at the risk of overload due to the number of criteria taken into account. She said her choice process had taken the shape of a funnel regarding the number of learning paths to compare, though not, unfortunately, regarding the number of criteria taken into account. Further quantitative research is required to investigate solutions aimed at reducing the risk of information overload by distinguishing between more and less important characteristics.

**Acknowledgments**

The work described here was sponsored by the TENCompetence Integrated Project, which has been funded by the European Commission’s 6th Framework Programme, priority IST/Technology Enhanced Learning. Contract 027087 (http://www.tencompetence.org).

**References**


AUPress: A Comparison of an Open Access University Press with Traditional Presses

Rory McGreal and Nian-Shing Chen
Technology Enhanced Knowledge Research Institute, Athabasca University, Canada // 1Department of Information Management, National Sun Yat-sen University, Taiwan // rory@athabascau.ca // nschen@mis.nsysu.edu.tw

ABSTRACT
This study is a comparison of AUPress with three other traditional (non-open access) Canadian university presses. The analysis is based on the rankings that are correlated with book sales on Amazon.com and Amazon.ca. Statistical methods include the sampling of the sales ranking of randomly selected books from each press. The results of one-way ANOVA analyses show that there is no significant difference in the ranking of printed books sold by AUPress in comparison with traditional university presses. However, AUPress, can demonstrate a significantly larger readership for its books as evidenced by the number of downloads of the open electronic versions.

Keywords
Open access, University presses, Printed books, Readership

Introduction
This investigation compares Canada’s first open access press, namely, AUPress with three other traditional Canadian university presses, which do not support open access at this time and whose editions are available only for purchase mostly in print format. AUPress, on the other hand, allows free downloading of its online edition under a Creative Commons, (Attribution-Noncommercial-No Derivative Works 2.5 Canada) licence and sells copies of its print editions. Open access is a model for scholarly publishing in which authors and publishers make their content freely available online with no requirements for authentication or payment. This analysis is based on actual physical book sales rankings on the largest online book retailer: Amazon.com and the Canadian version: Amazon.ca. Statistical methods are used to determine whether or not the traditional university presses show higher or lower sales rankings than the open press. This includes the sampling of the sales ranking of eleven randomly selected recently released books from each press on two occasions separated by three months in 2010 and one occasion a year later in 2011.

AUPress is Canada’s first fully open academic publisher. Founded in 2007, Athabasca University Press (http://www.aupress.ca/) released its first book in 2008. As of June, 2011, AUPress had published more than 50 books, which it produces in both electronic and print editions. Standard PDF format is used for both. In addition, the Press has acquired seven open access scholarly journals and two scholarly websites. The AUPress Editorial Committee, along with external peer reviewers drawn from a wide range of academic institutions, together serve as an assurance that the AUPress publications maintain a high level of quality.

One goal of the AUPress is to redress a situation in which scholars in developing countries write books or journal articles for publication in the West that, for reasons of cost, neither their colleagues at home nor their own university’s library will subsequently be able to access. It subscribes to the view that knowledge is not a commodity but should be freely accessible to all. Everything it produces is available on its website in the form of PDFs that can be downloaded, at no cost, by anyone who has an Internet connection, including classroom and e-learning uses. The books are licensed through the Creative Commons, giving anyone the right to reproduce the content for noncommercial purposes, as long as they attribute the source and make no derivative works. This thereby overcomes the most of the restrictions posed by copyright. More recently, the University of Ottawa Press has announced an open access collection, so this could be the beginning of a trend among Canadian universities in support of open access (University of Ottawa, 2010).

Athabasca University, of which the AUPress is part, is an open university, dedicated to overcoming barriers to education. In keeping with the university’s mission, AUPress cultivates research that pertains to distance learning and new educational technologies. It also focuses on publishing innovative and experimental works (including poetry and creative nonfiction) that challenge established ideas.
Literature survey

The OECD Centre for Educational Research and Innovation (CERI) has led in the promotion of open resources. It has identified several advantages supporting the use of open access materials. These include expanded access especially for non-traditional learners and promoting lifelong learning, bridging the gap between non-formal, informal and formal learning. The survey reports that institutions benefit by leveraging funding through free sharing and reuse of resources and improved quality (OECD, 2007). In addition, the IMS Global Learning Consortium has entered into a partnership with the Global Learning Resource Connection to enhance the discovery of educational resources, leveraging web technology to enable open electronic publishing (IMS Consortium, 2010).

Other open access proponents like Downes (2007, 2011) and Wiley (2007) stress the need for sustainability of open materials in terms of funding, technology and content. Downes describes several different such sustainability models that are presently in use: Memberships, Donations, Conversions, Contributor Pays, Sponsorships, Institutional, and Governmental. Wiley agrees and also adds other models such as the Sakai Platform model where universities collaborate and share resources, and the Volunteer model.

In addition, Antelman (2004) has revealed some evidence that open access articles have a greater research impact, primarily because of their increased accessibility. This greater impact factor enjoyed by open access publications is supported by research conducted by Harnad and Brody (2004), who claim that OA dramatically increases the number of potential users of any given article.

Nicholas, Rowlands, Clark, Huntington, et al. (2008) conducted “the biggest survey of its kind” on the use of scholarly ebooks. Surprisingly, the investigation showed that there was already a major penetration of ebooks in academia. Their study discovered that academics and students of every age group do read online “with alacrity” and they do not read whole books, but rather bits of them. This rather explains the propensity of the majority of readers accessing the AUPress website, who choose chapters rather than whole books for download.

This survey was followed up by a JISC Observatory (2009) study that determined that ebooks had no conclusive negative impact on UK print sales. Moreover, Hilton III and Wiley (2010), in their comparison of free ebooks and traditional sellers, saw sales gains after the free versions were released. This was in contrast to their initial hypothesis that book sales would decline. Their research demonstrated that a correlation exists between a free ebook and increased print sales. This study on Amazon rankings was designed to provide complementary data to either support or refute these hypotheses.

In addition, Eysenbach (2006) concluded that open access articles are more immediately recognized and cited by peers than non-open access articles. This finding was supported by Donovan and Watson (2011), who confirmed in their study that open access improves an article’s research impact noting that open access articles “can expect to receive 50% more citations than non-open access writings of similar age from the same venue.”

Amazon sales ranking

The Amazon sales ranking number is provided as a service for authors and publishers and can be one reasonable gauge of the number of printed books purchased. The ranking provides a relative measure that is useful for assessing a book’s sales performance on Amazon. The lower ranking number of a particular book can be interpreted as signifying higher sales. Two ranking lists were studied, based on both Amazon.com and Amazon.ca sales, which are updated each hour to reflect recent and historical sales of every book sold on the respective web sites. For competitive reasons, Amazon does not release actual sales information to the public, so very few, if any people outside of Amazon know the actual sales numbers (Amazon, 2010).

Significantly, this rating does not apply to Kindle books that have been increasing rapidly in sales volume (Rosenthal, 2010). The introduction of the Kindle, iPad and other ebook readers and tablets are beginning to have an enormous impact on electronic and print book sales. Amazon is now selling more electronic than print books (Allen, 2009; TechBuzz, 2010). As the electronic publishing industry matures, it will be increasingly important to research the effects of the free distribution of electronic books. Meanwhile, their rankings refer to printed book sales only.
Rampant Tech Press (n.d.) and Sampson (2010) have independently ventured to extrapolate the sales to a ranking order. They have come up with similar information displayed on the following table (Table 1).

<table>
<thead>
<tr>
<th>Rank #</th>
<th>Rampant Press Copies Sold/day</th>
<th>Sampson copies per week</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; #1</td>
<td>3000</td>
<td>&gt; 1,000 copies per week</td>
</tr>
<tr>
<td>&gt; #10</td>
<td>650</td>
<td>200 – 1,000 copies per week</td>
</tr>
<tr>
<td>&gt; #100</td>
<td>100</td>
<td>100 – 200 copies per week</td>
</tr>
<tr>
<td>&gt; #1000</td>
<td>13</td>
<td>10 – 100 copies per week</td>
</tr>
<tr>
<td>&gt; #10,000</td>
<td>2.2 (11 copies every 5 days)</td>
<td>1 – 10 copies per week</td>
</tr>
<tr>
<td>&gt; #100,000</td>
<td>0.2 (1 copy every 5 days)</td>
<td>&lt; 200 books sold</td>
</tr>
<tr>
<td>&gt; #1,000,000</td>
<td>0.006 (3 copies every 500 days)</td>
<td>&lt; 40 books sold</td>
</tr>
<tr>
<td>&gt; #2,000,000</td>
<td>0.0001 (1 copy every 1000 days)</td>
<td>1 book ordered</td>
</tr>
</tbody>
</table>

Rosenthal (2010) also provides similar estimates, noting that the lower ranking books (those with a higher ranking number, (>100,000) move comparatively little in their ranking as opposed to rather erratic movements in the best sellers (<10,000). He notes that weak sellers decay relatively slowly. He observes that a title must sell at least one copy a year to remain above a rank of two million. As most academic books never reach the lowest rankings, they are with a few exceptions, to be considered “weak sellers” (>100,000).

Sampson (2010) notes that the Amazon rankings provide only marginal sales data that are rough estimates at best. On the other hand he claims that the relative sales ranking can be useful for comparisons among books. Books with rankings between #10,000 and #100,000 are recalculated once a day; historic sales information plays a key role in these calculations. However, with books ranking higher than #100,000, which are also recalculated every day, history takes a back seat.

**Methodology**

Stratified sampling is a common probability method that is considered to be better than random sampling because the stratification reduces sampling error. The relevant stratum in this case was a subgroup of books published between 2008 and 2010. This was necessary because the targeted population consisted of AUPress books. As AUPress is new, it only had published books in those years. Random sampling was then used to select a reasonable number of samples (n=11) from each publisher. This provided the researchers with confidence that the strata represented each population reasonably well and accurately represented the overall publications in the years under investigation. Limiting the other presses to a subgroup made up of their most recent books published ensured a fair comparison with the new AUPress.

The sampled publications were then investigated to determine their ranking order on both Amazon.com and Amazon.ca. It was considered appropriate to investigate both “stores” as it was expected that Canadian scholarly publications would be relatively better sellers in Canada than internationally. The survey was also conducted on three dates separated, the first two separated by three months and the last by one year (January and April, 2010 and April, 2011). This date separation is recommended to get more trustworthy ranking numbers as the numbers can be skewed drastically if measured on any one occasion (Rosenthal, 2010; Shepard, 2010; & Sampson, 2010).

**The investigation**

The investigation aimed to determine whether or not there was a ranking difference between the ranking of the books in the open press and any of the traditional presses. AUPress (AUP) which is the open access university press was compared to the following three traditional presses: University of Toronto Press (UTP), University of Calgary Press (UCP) and University of Alberta Press (UAP) in terms of sales ranking of these presses from Amazon (amazon.ca & amazon.com). The aim of the study was to determine if there is a significant difference between the open access press and the traditional presses using the mean sales rank.
The *Null Hypothesis* was posited that the mean sales ranks of AUP, UTP, UCP and UAP are equal. This was tested by the ANOVA analysis against the *Alternative Hypothesis* that the mean sales ranks of AUP, UTP, UCP and UAP are not equal.

The Amazon.com and Amazon.ca ranking results for these four university presses are available in the following Tables 2 and 3 for January, 2010; and Tables 4 and 5 for April, 2010.

### Table 2. Rankings from Amazon.ca, January 2010

<table>
<thead>
<tr>
<th>Athabasca University Press</th>
<th>University of Toronto Press</th>
<th>University of Calgary Press</th>
<th>University of Alberta Press</th>
</tr>
</thead>
<tbody>
<tr>
<td>57,105</td>
<td>227,397</td>
<td>422,660</td>
<td>154,521</td>
</tr>
<tr>
<td>198,141</td>
<td>119,746</td>
<td>111,002</td>
<td>355,812</td>
</tr>
<tr>
<td>239,621</td>
<td>46,419</td>
<td>396,751</td>
<td>424,099</td>
</tr>
<tr>
<td>98,969</td>
<td>56,934</td>
<td>561,944</td>
<td>246,631</td>
</tr>
<tr>
<td>101,707</td>
<td>201,532</td>
<td>683,365</td>
<td>169,208</td>
</tr>
<tr>
<td>225,921</td>
<td>227,397</td>
<td>1,195,769</td>
<td>65,710</td>
</tr>
<tr>
<td>145,839</td>
<td>249,305</td>
<td>237,886</td>
<td>60,384</td>
</tr>
<tr>
<td>488,360</td>
<td>477,072</td>
<td>421,807</td>
<td>83,253</td>
</tr>
<tr>
<td>80,031</td>
<td>283,831</td>
<td>270,707</td>
<td>91,869</td>
</tr>
<tr>
<td>408,713</td>
<td>419,100</td>
<td>388,270</td>
<td>267,048</td>
</tr>
<tr>
<td>122,315</td>
<td>332,398</td>
<td>787,757</td>
<td>197,166</td>
</tr>
</tbody>
</table>

### Table 3. Rankings from Amazon.com, January 2010

<table>
<thead>
<tr>
<th>Athabasca University Press</th>
<th>University of Toronto Press</th>
<th>University of Calgary Press</th>
<th>University of Alberta Press</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,260,279</td>
<td>2,393,121</td>
<td>3,124,635</td>
<td>1,290,317</td>
</tr>
<tr>
<td>705,438</td>
<td>3,337,710</td>
<td>160,272</td>
<td>3,428,847</td>
</tr>
<tr>
<td>1,062,251</td>
<td>1,190,429</td>
<td>1,048,357</td>
<td>4,068,647</td>
</tr>
<tr>
<td>1,765,283</td>
<td>735,372</td>
<td>1,797,624</td>
<td>776,928</td>
</tr>
<tr>
<td>2,940,755</td>
<td>2,992,991</td>
<td>647,557</td>
<td>1,365,207</td>
</tr>
<tr>
<td>4,472,042</td>
<td>2,393,121</td>
<td>3,076,338</td>
<td>999,705</td>
</tr>
<tr>
<td>1,086,172</td>
<td>1,483,875</td>
<td>724,521</td>
<td>334,671</td>
</tr>
<tr>
<td>1,712,101</td>
<td>2,376,571</td>
<td>4,938,289</td>
<td>2,865,188</td>
</tr>
<tr>
<td>2,637,674</td>
<td>2,248,576</td>
<td>4,312,491</td>
<td>4,205,723</td>
</tr>
<tr>
<td>2,087,648</td>
<td>618,051</td>
<td>3,634,196</td>
<td>8,581,611</td>
</tr>
<tr>
<td>1,068,800</td>
<td>1,654,718</td>
<td>2,006,625</td>
<td>3,419,384</td>
</tr>
</tbody>
</table>

### Table 4. Rankings from Amazon.ca, April 2010

<table>
<thead>
<tr>
<th>Athabasca University Press</th>
<th>University of Toronto Press</th>
<th>University of Calgary Press</th>
<th>University of Alberta Press</th>
</tr>
</thead>
<tbody>
<tr>
<td>168,692</td>
<td>123,953</td>
<td>211,059</td>
<td>518,411</td>
</tr>
<tr>
<td>198,349</td>
<td>147,143</td>
<td>292,372</td>
<td>285,057</td>
</tr>
<tr>
<td>265,556</td>
<td>210,104</td>
<td>151,690</td>
<td>456,564</td>
</tr>
<tr>
<td>157,685</td>
<td>96,042</td>
<td>616,459</td>
<td>262,456</td>
</tr>
<tr>
<td>168,185</td>
<td>278,407</td>
<td>762,310</td>
<td>503,667</td>
</tr>
<tr>
<td>161,030</td>
<td>123,953</td>
<td>464,457</td>
<td>190,884</td>
</tr>
<tr>
<td>381,654</td>
<td>312,514</td>
<td>240,932</td>
<td>8,851</td>
</tr>
<tr>
<td>530,916</td>
<td>233,513</td>
<td>454,194</td>
<td>230,104</td>
</tr>
<tr>
<td>314,160</td>
<td>400,998</td>
<td>88,945</td>
<td>207,242</td>
</tr>
<tr>
<td>437,952</td>
<td>216,795</td>
<td>349,672</td>
<td>155,859</td>
</tr>
<tr>
<td>130,555</td>
<td>351,933</td>
<td>887,941</td>
<td>183,728</td>
</tr>
</tbody>
</table>
Table 5. Rankings from Amazon.com April 2010

<table>
<thead>
<tr>
<th>Presses</th>
<th>Athabasca University Press</th>
<th>University of Toronto Press</th>
<th>University of Calgary Press</th>
<th>University of Alberta Press</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2,118,462</td>
<td>2,604,504</td>
<td>3,351,468</td>
<td>338,028</td>
</tr>
<tr>
<td></td>
<td>2,857,528</td>
<td>590,544</td>
<td>489,905</td>
<td>4,967,211</td>
</tr>
<tr>
<td></td>
<td>1,079,861</td>
<td>1,165,025</td>
<td>474,650</td>
<td>4,378,205</td>
</tr>
<tr>
<td></td>
<td>184,133</td>
<td>723,673</td>
<td>851,699</td>
<td>695,601</td>
</tr>
<tr>
<td></td>
<td>4,945,041</td>
<td>2,887,859</td>
<td>2,498,913</td>
<td>200,399</td>
</tr>
<tr>
<td></td>
<td>4,871,260</td>
<td>2,604,504</td>
<td>759,218</td>
<td>1,949,501</td>
</tr>
<tr>
<td></td>
<td>1,432,406</td>
<td>1,498,134</td>
<td>888,175</td>
<td>1,008,402</td>
</tr>
<tr>
<td></td>
<td>1,754,331</td>
<td>2,752,047</td>
<td>1,208,738</td>
<td>3,069,183</td>
</tr>
<tr>
<td></td>
<td>2,855,897</td>
<td>2,398,991</td>
<td>4,647,516</td>
<td>4,525,754</td>
</tr>
<tr>
<td></td>
<td>1,236,135</td>
<td>855,558</td>
<td>3,918,130</td>
<td>4,067,474</td>
</tr>
<tr>
<td></td>
<td>664,396</td>
<td>1,696,327</td>
<td>1,989,543</td>
<td>3,672,513</td>
</tr>
</tbody>
</table>

By merging the rankings of Table 2 to Table 5, a total of 44 data sets for each of the press were used for the data analysis. Table 6 shows the mean, standard deviation and standard error of all the four university presses. One-way ANOVA was then used to test if there is any significant difference among these four presses. The result (Table 7) shows there is no significant difference $F(3,172) = .761$, $p = 0.518$, therefore the Null Hypothesis cannot be rejected. This implies that academic books on open access do not lessen printed book sales online in comparison with traditional university presses.

Table 6. The Mean, Std Deviation and Std. Error

<table>
<thead>
<tr>
<th>Presses</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUP</td>
<td>44</td>
<td>1133622</td>
<td>1310647</td>
<td>197587</td>
<td>57105</td>
<td>4945041</td>
</tr>
<tr>
<td>UTP</td>
<td>44</td>
<td>1053141</td>
<td>1029137</td>
<td>155148</td>
<td>46419</td>
<td>3337710</td>
</tr>
<tr>
<td>UCP</td>
<td>44</td>
<td>1285155</td>
<td>1369355</td>
<td>206438</td>
<td>88945</td>
<td>4938289</td>
</tr>
<tr>
<td>UAP</td>
<td>44</td>
<td>1484705</td>
<td>1920836</td>
<td>289577</td>
<td>8851</td>
<td>8581611</td>
</tr>
<tr>
<td>Average</td>
<td>44</td>
<td>1239156</td>
<td>1407494</td>
<td>212188</td>
<td>50330</td>
<td>5450663</td>
</tr>
</tbody>
</table>


Table 7. One-way ANOVA Analysis Results

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>$4.759e^{12}$</td>
<td>3</td>
<td>$1.586e^{12}$</td>
<td>.761</td>
<td>.518</td>
</tr>
<tr>
<td>Within Groups</td>
<td>$3.587e^{14}$</td>
<td>172</td>
<td>$2.085e^{12}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$3.635e^{14}$</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUPress book downloads

In the six months prior to this survey first being conducted, there were a total of more than 24,000 chapter downloads, of which more than 9,000 were full books. The average total number of downloads was 2,000 and the full book average was over 700. The median download rate for full books was 226 and the total downloads median was 798 (including chapters). Some of the more popular scholarly books, particularly those in the elearning field, had more than 2,000 full book downloads and over 6,000 total downloads (including chapters) as shown in Table 8.

Table 8. Six Month Book + chapter downloads at AUPress

<table>
<thead>
<tr>
<th>Books</th>
<th>Aug'09</th>
<th>Sep'09</th>
<th>Oct'09</th>
<th>Nov'09</th>
<th>Dec'09</th>
<th>Jan'10</th>
<th>Total</th>
<th>Full Books</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>98</td>
<td>105</td>
<td>166</td>
<td>193</td>
<td>117</td>
<td>119</td>
<td>798</td>
<td>333</td>
</tr>
<tr>
<td>B</td>
<td>73</td>
<td>55</td>
<td>75</td>
<td>51</td>
<td>86</td>
<td>76</td>
<td>416</td>
<td>38</td>
</tr>
<tr>
<td>C</td>
<td>93</td>
<td>90</td>
<td>141</td>
<td>114</td>
<td>75</td>
<td>94</td>
<td>607</td>
<td>69</td>
</tr>
</tbody>
</table>
- all requests made by known robots and spiders have been excluded from the download counts
- If multiple requests come from the same visitor in a relatively short period of time (a few hours) FOR THE SAME FILE, the requested is counted as a single request

AUPress books and chapters have been downloaded by scholars and other users in more than sixty different countries. As expected the largest numbers of downloaders (more than 50%) are from Canada and the USA, but more than 33% of the other downloaders were from developing countries (2314). Others were from the emerging countries of eastern Europe (385). Several books have also won distinguished international academic awards and have been reviewed and cited in leading scholarly journals. So, open access scholarly publications can claim a much higher and more global readership than traditional publications. The Amazon ranking data suggests that, at least in one measure, open access books sell as well as traditional press books, and the large number of downloads of open access books point to a significant advantage over traditional publications in terms of total number of readers.

Conclusion

Results suggest that there is no significant difference in the Amazon rankings. This suggests that releasing academic books on open access does not lessen printed book sales online in comparison with traditional university presses using Amazon.com and Amazon.ca rankings. On the other hand, AUPress, because it is open access and publicly available at no cost, can boast of having a significantly larger readership for its books. The traditional university presses, because of their cost, print-only format, and other proprietary limitations are not readily available and therefore not accessible to many potential readers.

As has been noted, the introduction of the Kindle, iPad and other eBooks and tablets are beginning to have an enormous impact on electronic and print book sales. As the electronic publishing industry matures it will be increasingly important to research the effects of the free distribution of electronic books. Nevertheless, the results of this investigation must be viewed with some caution. These results cannot be easily generalized to other book sales. Causation has not been proven. In addition, the wide differences among the rankings of individual books were not factored into this study. As more open access presses are established, a larger sampling pool should be attempted to determine whether or not there is a more robust relationship between relative ranking of the different books and the impact of open access publishing.

Publishers are very reluctant to provide book sale numbers to researchers, so unfortunately this study could not address other sales. So, we were not able to determine actual sales, nor could we factor in the sales to university libraries, which are major purchasers of academic books. However, this is changing as Amazon and other online booksellers are becoming more convenient sources for purchases. Future analysis needs to be done to determine whether open access total book sales and library book sales are significantly different.

Traditional publishers have argued that open access publication will inevitably undercut sales of printed books and is thus inherently unsustainable. The experience of AU Press suggests otherwise. Like all university presses, AU Press receives funding from its parent university. But, to support its open access program, the Press also earns revenue from the sale of its books, both in traditional print format and as enhanced electronic files (Universal PDFs or epub files). The latter are currently purchased chiefly by libraries, which, for reasons of convenience, prefer to acquire
new titles through ebook vendors who represent a wide array of presses. Even when a book’s content is available for free online, many individual readers still choose to purchase print. In addition, online publications are more readily discoverable, which increases the number of potential customers and thus serves as a marketing tool.

Moreover, there is the significant added advantage as shown in this paper, namely substantially increasing readership, even in developing countries, of scholarly books that are made available on line. To date, the AUPress’s books have been accessed more than 120,000 times by learners. This aids scholars in their obligation to disseminate their research and can contribute significantly to their citation by other researchers who have open access to their ideas.

Although this study has focused on the Amazon ranking related to print sales, institutions could also benefit in other ways unrelated to actual sales. For example, the wider dissemination of knowledge not only serves to extend a university’s research mandate, but it also provides positive publicity among learners around the world. This can be very valuable in enhancing the university’s reputation and possibly increasing enrollments.

References


## Appendix

### AUPress Book Description

<table>
<thead>
<tr>
<th>Author/Editor</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Brink</td>
<td><em>Imagining Head Smashed In</em></td>
</tr>
<tr>
<td>B Karras</td>
<td><em>Northern Rover</em></td>
</tr>
<tr>
<td>C Carter</td>
<td><em>Importance of Being Monogamous</em></td>
</tr>
<tr>
<td>D Jameson</td>
<td><em>One Step over the Line</em></td>
</tr>
<tr>
<td>E Anderson</td>
<td><em>Theory and Practice of Online Learning</em></td>
</tr>
<tr>
<td>F Atkinson</td>
<td><em>Making Game</em></td>
</tr>
<tr>
<td>G Foran</td>
<td><em>Expansive Discourses</em></td>
</tr>
<tr>
<td>H Ally</td>
<td><em>Mobile Learning</em></td>
</tr>
<tr>
<td>I Smith</td>
<td><em>Liberalism Surveillance and Resistance</em></td>
</tr>
<tr>
<td>J Perry</td>
<td><em>More Moments in Time</em></td>
</tr>
<tr>
<td>K Allan</td>
<td><em>Bomb Canada</em></td>
</tr>
<tr>
<td>L MacDonald</td>
<td><em>Beaver Hills</em></td>
</tr>
<tr>
<td>M Power</td>
<td><em>Designer’s Log</em></td>
</tr>
</tbody>
</table>
Building Lectures and Building Bridges with Socio-economically Disadvantaged Students

Peter Phillips and Birgit Loch

School of Accounting, Economics and Finance, University of Southern Queensland, Toowoomba, QLD 4350, Australia // ¹Faculty of Engineering and Industrial Science, Swinburne University of Technology, Hawthorn, VIC 3122, Australia // phillipp@usq.edu.au // Birgit.Loch@swin.edu.au

ABSTRACT

This paper is an empirical analysis of the first stage of an ongoing effort to introduce technology to enhance student learning in introductory corporate finance within a multi-campus and multi-mode regional Australian University. The engagement and performance of low socio-economic status (SES) students is of particular interest because approximately one-quarter of the university’s enrolled students are classified low SES. A Tablet PC is used to facilitate a cooperative ‘building’ of each week’s lecture in class and the recording of this process for delivery online. The analysis of the academic achievement of two cohorts of students in two different semesters—with the technology and without—forms the basis of the formal evaluation of the efficacy of the approach to date. The results indicate that there is a significant difference in retention (drop-out statistics) and academic achievement (examination performance, final grade and course progression statistics) between the ‘Tablet PC’ and ‘control’ semesters. The largest improvement was exhibited by the low SES students.

Keywords
Tablet PC, Socio-economic status, Academic achievement, Retention, Student engagement

Introduction

This paper presents the results of an analysis of the performance of two cohorts (2008 and 2009) of students in the introductory corporate finance course at the University of Southern Queensland (USQ). In 2008, a traditional approach was taken to the delivery of the course in both face-to-face and distance modes. In 2009, the delivery of lectures was undertaken with the assistance of a Tablet PC and the (live) lecture recording ‘screencasts’ produced using the technology were made available to all students.

The present study investigates the efficacy of teaching technologies by exploring the differences in student engagement and performance during the Tablet PC semester against the control or non-Tablet PC semester. However, rather than investigate these factors in isolation, this study focuses on the engagement and performance within the added dimension low, medium and high socio-economic status (SES).

The educational psychology literature identifies a relationship between SES and academic achievement. The interaction of variables and the ways in which SES may influence (directly and indirectly) academic achievement help to explain our results and illuminate some important possible paths for future research. However, it should be noted that the objective of this investigation is not to examine and identify a relationship between SES and academic achievement (which is the focus of much of the educational psychology literature) but rather to highlight some possible benefits to the use of teaching technologies within a context in which the students enrolled in a university course are characterised by a diversity of SES. Specific investigation of the association between SES, achievement and teaching technology deployment is a broader research program that requires careful consideration of particular points of measurement and methodology.

A research program that focuses on aspects of delivery of tertiary level education to low SES students is important for several reasons. Of course, there are many social justice and equity issues that can be brought to bear in justifying special focus on low SES students. In Australia, contemporary government policy intensifies the need for research and action with regard to tertiary education and low SES students. The Australian Government’s (2009) ‘Transforming Australia’s Higher Education System’ policy document, which was drafted in response to the Bradley Review of Higher Education in Australia (Bradley et al. 2008), places significant emphasis on participation in higher education of people from low SES backgrounds. More than AU$400 million of government funding has been allocated to support participation targets with the ultimate goal of having 40 percent of 25 to 34 year olds bachelor degree (or above) qualified by 2025 and low SES students constituting 20 percent of undergraduate enrolments by 2020 (Commonwealth Government 2009). The challenges facing the Australian university system are substantial.
The role that technology can play in helping to overcome these challenges is an important research program for this reason. This paper represents some very preliminary steps and generates conclusions from a higher education institutional context where, already, more than 20 percent of students are from low SES backgrounds.

This paper is organised as follows. First, an overview of the literature is presented with focus on those studies that deal with the relationship between SES and academic achievement. Then, the approach taken to the delivery of introductory level corporate finance is described and the pivotal role played by the Tablet technology is outlined. The data is analysed after that, highlighting a substantial improvement in student performance during the Tablet PC semester mainly from low and medium SES backgrounds. Results are discussed and some directions for future research are presented.

**Literature overview**

There are two main streams of literature that are relevant to the background context of the present study. First and foremost, the ‘educational psychology’ literature examines the effect of SES on academic achievement directly or indirectly. These studies investigate sociological-psychological aspects of education, development and achievement, including the relationship between SES and self-efficacy (and achievement). Second, in the ‘economics and sociology of education’ literature, various aspects of education and education policy are examined from an economic and sociological perspective. These investigations examine the relationships between various SES-related variables and variables such as drop-out rates, choice of subjects and careers. Together, these two streams of research examine the complicated fabric in which SES and other aspects of education and student behaviour are intertwined.

**Educational Psychology**

The investigation of the relationship between SES and academic achievement has been explored on many occasions. The meta-analyses or reviews of the literature undertaken by White (1982) and Sirin (2005) highlight the extensive nature of the literature in this field of study. A relationship exists between SES and academic achievement but the relationship is sensitive to a variety of measurement and methodological factors. As Sirin (2005, p.438) explains, “…methodological characteristics, such as the type of SES measure, and student characteristics, such as student’s grade, minority status, and school location, moderated the magnitude of the relationship between SES and academic achievement.” The relationship is still clear and strong enough, however, to permit statements such as the following: “Socio-economic status differences in children’s reading and educational outcomes are ubiquitous, stubbornly persistent and well documented” (Aikens and Barbarin, 2008, p.235). The relationship between SES and academic achievement is due to a complex interaction of a number of variables, it appears to be generally accepted that SES impacts to a considerable extent on various aspects of students’ learning experiences.

Some of the most important insights generated by the extant investigations into this relationship are the ways in which SES affects academic achievement. In addition to measurement problems, the underlying relationship is a complex one to investigate (even if the other factors were absent). This is because the effects of SES are ‘channelled' through family, neighbourhood and school contexts (Aikens and Barbarin 2008). These contexts are shaped by factors such as access to books and other resources (Evans, 2004; Duncan, Yeung, Brooks-Gunn and Smith, 1998; Whitehurst and Lonigan, 1998; Duncan, Brooks-Gunn and Klebanov, 1994); various familial interaction variables, including ‘negative parenting’, violence, separation and disruption (Evans, 2004; Raviv, Kessenich and Morrison, 2004; Emery and Laumann-Billings, 1998); and numerous extra-familial variables, including crime, low quality social networks, aggressive or disadvantaged peers, and low school and teacher quality (Evans, 2004; Bandura, Barbaranelli, Caprara and Pastorelli, 2001; Sampson, Raudenbush and Earls, 1997; Bandura et al. 1996; Federman, Garner, Short, Cutter, Levine, McGough and McMillen, 1996). SES influences the risk of exposure to all of these factors. The complex interaction and cumulative effects of these variables affects the academic achievement of students. This leaves aside the effect that SES can have on development (and, indirectly, on academic achievement) through its influence on physical and psychological health (Bradley and Corwyn, 2002).

By increasing the risks of exposure to crime, violence, disruption and other similar factors, it is easy to understand how SES might affect academic achievement. There are, however, many ways in which SES might indirectly affect academic achievement and overcoming the challenges presented by SES necessitates an understanding of these less
obvious relationships. Caprara, Fida, Vecchione, Del Bove, Vecchio, Barbaranelli and Bandura (2008, p.527) state, “Diverse lines of evidence show that SES affects performance, in large part, through its impact on psychosocial processes rather than directly.” For example, Miech, Essex and Goldsmith (2001) found that self-regulation, which refers to processes, such as the tendency to maintain attention on a task and to suppress inappropriate behaviour under instructions, serves as a moderator of the relationship between SES and adjustment to school. Bandura et al. (2001) examine the interaction of socio-economic conditions and other factors such as educational aspirations, perceived efficacy and scholastic accomplishments in contributing to career choice and development (see Bandura et al. 2001, p.200). The indirect influence of SES factors on academic achievement led Caprara et al. (2008) to control for SES in their investigation of self-regulatory efficacy and academic performance. Self-regulatory efficacy refers to a student’s judgements of capability to perform successfully at designated levels (Schunk, 1991) and students’ beliefs in their efficacy to regulate their learning activities and master academic subjects (Caprara et al. 2008, p.525). It plays a central role in academic development, and, thereby academic achievement, may be impacted upon by SES factors. It is necessary, therefore, to be aware of the ways in which indirect influences may be addressed. For example, particular delivery approaches within a university context may help to overcome or offset the negative impact of SES on variables such as self-regulatory efficacy.

The economics and sociology of education

Whereas the investigation of the relationship between SES and development, academic achievement, motivation and engagement has been undertaken by educational psychologists, another part of the important story concerning the relationship between SES and student choices and behaviour has been contributed to by sociologists and, to a certain extent, economists. This literature considers the effect of SES or related concepts on aspects of student choice and behaviour without focusing heavily on cognitive psychological variables or constructs. A large proportion of this research deals with the relationship between aspects of student choice and behaviour and economic resources or related factors (Bozick, 2007; Entwisle, Alexander and Olson, 2005; Marsh and Kleitman, 2005; Christie, Munro and Fisher, 2004; Werfhorst, Sullivan and Cheung, 2003). Whilst some of this literature overlaps with the contributions listed in the previous sub-section to the extent that the ways in which SES influences student academic achievement and engagement are addressed, the sociological components of some studies distinguishes them from those listed previously. For example, De Graaf, De Graaf and Kraaykamp (2000) investigated the role of financial and cultural resources (cultural capital) on students’ educational attainment. Most recently, Meeuwisse, Severiens and Born (2010) examined the interaction of multiple variables in students’ decisions to withdraw from higher education. They support the general theme that emerges in all of the studies reviewed herein: The interplay of variables that characterises the investigation of SES and aspects of students’ behaviour, choices and outcomes is tremendously complex.

The complex interaction of SES and related factors and student achievement, motivation and engagement sits in the background of attempts by governments to frame and execute sound education policy and, more broadly, macroeconomic policy. The economics of education, which can be traced as a research program to the early 1960s (particularly, Schultz, 1961), encompasses investigations into the allocation and financing of education spending (Fernandez and Rogerson, 2003; Hoxby, 2001), the impact of education on the economic system, the economic impact of schools, colleges and universities on local communities (Siegfried, Sanderson, McHenry 2007), optimisation, allocation and choice problems (Ladd, 2002; Card and Krueger, 1996) and various points of theory that underlie and focus empirical investigations, including the status of education as investment or consumption and debates concerning human capital and screening (see Schultz, 1961; Blaug, 1968; Tramonte and Willms, 2009). Economists have also investigated various aspects of SES and education, including issues such as wage discrepancies between different groups (for example, O’Gorman, 2009 and Clotfelter, 1999, esp. p.8). Much of this work, particularly the now-classical contributions to the economics of education, has shaped the educational policies of many Western governments over the past five decades. It is easy to overlook the complexities that underlie macroeconomic educational policy.

The macroeconomics of education takes place against a backdrop of the actions of individual students and institutions. The Australian government’s higher education policy is based on the premise that, “[h]igher education fuels economic development, productivity, high skilled jobs and supports Australia’s role as a middle power and leader in the region” (Commonwealth Government, 2009). As explained earlier, a key feature of this policy is the wide participation in higher education by people from low SES backgrounds. The policy, which aims to ‘support
high quality teaching and learning and improve access and outcomes for students from low socio-economic backgrounds’, necessitates effective utilisation of resources by universities in order to achieve access, outcomes and quality-related objectives.

The implications of the literature surveyed in this section may be summarised as follows. First, the effect of SES on student achievement, motivation, engagement and choice, is not clear-cut but a component of a series of complex interactions among many interrelated variables. Second, the direct, indirect and cumulative effects of SES are significant and must be considered important by institutions of higher education, particularly those whose student population consists of diverse backgrounds and higher proportions of low SES students. Third, students’ choices, such as the choice to drop-out or the necessity of part-time work, and students’ self-regulation and self-regulatory efficacy are influenced by SES or related factors. The present investigation presents the results of a preliminary trial of educational technology and an assessment of its effect on student performance and choice to drop-out. The context is a core course delivered within a university that is characterised by a high percentage of low SES students and students studying by distance. The literature surveyed above implies that this combination may be particularly conducive to high drop-out rates and low performance. It is the purpose of this investigation to present the results of some very first steps towards assessing the role that educational technologies can play in offsetting or overcoming the disadvantages associated with SES-related factors.

Building lectures with a Tablet PC

In this section, the approach to delivering the introductory corporate finance course is outlined. Introductory corporate finance at USQ—a multi-campus and multi-mode institution—is offered bi-annually. In semester 1 (March to July), the average number of student enrolments is 400. Of these, approximately 100 attend classes at the Toowoomba campus, 40 attend classes at the Springfield campus, 20 attend classes at the Fraser Coast campus and approximately 150 study by distance. The remainder of the students study a combination of distance mode and face-to-face teaching (undertaken by a local tutor) at USQ’s off-shore partner institutions in China and Singapore. Classroom teaching in a lecture theatre is supplemented or complemented by communication via the learning management system (LMS).

In semester 1 2008, the introductory corporate finance course was characterised by standard face-to-face lecture and distance (online) delivery of study material. Lectures were ‘traditional’ dot-point PowerPoint presentations fully prepared and delivered during the lecture as a standard slideshow. No recording of the lectures was made available. In semester 1 2009, a trial of Tablet PC technology was undertaken. The Tablet technology facilitated a number of innovations in the nature of the classroom lecture delivery and, importantly, the lecture ‘screencasts’ were made available online for all students to view. In this approach, emphasis is placed upon the building or construction of the lecture together with the students in the lecture theatre using only a basic framework of PowerPoint slides for structure and as a writing environment. This much more interactive approach, which is facilitated effectively by the technology, is not only valued by the on-campus and distance students but also creates an environment that is much more rewarding for the instructor.

The main innovation facilitated by the Tablet PC was the interactive construction of the lectures in class. Whereas the lectures were once fully completed before class and delivered as a complete product, the 2009 PowerPoint slides that were taken to class only consisted of the structure and not much detail. Two typical examples from 2008 (standard) and 2009 (Tablet) are presented in Figure One. The standard PowerPoint slides contain most of the points that will be made during the lecture. The reason for this approach was to provide a ‘static record’ of the lecture for the students studying by distance. On the other hand, the Tablet PC PowerPoint slides contain, for the most part, only slide titles and, on occasions, pictures to prompt discussion. As the lecture progresses the slides are ‘inked’ using the Tablet PC and the completed lecture gradually takes shape. Each action that takes place (audio and the inking) is captured using Camtasia Studio and a ‘screencast’ is produced and uploaded to the course site on the LMS. All students are then able to download, watch and listen to the lecture. This provision of a classroom experience to students studying by distance by itself is, of course, a valuable innovation.
At the beginning of each lecture, students were provided with (paper) handouts of the slides that represented the building blocks for the lecture. Most of the slides had no more detail than the right-hand (non-standard) slide in Figure 1. The lecture proceeded with the instructor using the Tablet PC at the front of the theatre (always facing the students) and the Tablet PC screen and lecture slides visible to students on the large projection screen above the instructor. At each slide, the instructor could jot down some important points, equations, examples or diagrams (with audio fully recorded). Students completed or filled in their ‘blank’ handouts. On numerous occasions in any particular lecture, the instructor involved the students by calling for suggestions, explanations or results of particular calculations, and engaging students in discussion. In this way, the lectures were ‘built’ in class with the students and the construction was shaped and contributed to by their participation; see Figure 2 for an example.

With regards to teaching practice and issues such as engagement and interaction within a classroom environment, the approach described in this section involves considerably more student participation than the standard slide-show delivery. This may contribute to improvements in student performance and motivation in several ways, including the generation of a more positive classroom environment (Phillips and Loch, 2010). The online availability of the screencasts provides all students the opportunity to engage with the course at a much higher level and experience a number of important elements of the classroom experience. In the next section, student performance in semester 1 2008 is compared with performance in semester 1 2009. Emphasis is placed upon the performance of students designated low SES.

**Student performance**

The impact of the Tablet PC deployment and the approach to lecture delivery described in the previous section is assessed by comparing the performance of students in the non-tablet semester in 2008 with the performance of students in the tablet semester in 2009. During these two semesters, the course team, including the principal
instructor, and type and level of difficulty of assessment, remained unchanged. The approach to delivering the lectures at the main Toowoomba campus (and the availability of the screencasts) is the only significant difference between the two semesters. The educational psychology literature identifies many factors that contribute to student motivation, engagement and performance (see, for example, Miller, Greene, Montalvo, Ravindran and Nichols, 1996). It is likely that any improvement in these areas is the result of a complex interaction of a number of variables. The approach taken here in comparing performance across a ‘control’ semester and a semester within which a teaching innovation is trialed, mirrors that which has been taken in previous studies (see Sosin, Blecha, Agarwal, Bartlett and Daniel, 2004; Ball, Eckel and Rojas, 2006; Salemi, 2009). In the following sub-section, a summary is given of the formal analysis of failing grade distribution from (Phillips and Loch, 2010) across all campuses. The difference in exam performance is also shown, to provide a context for the main topic of this paper: evaluation of the impact of technology use on motivation and performance by socio-economic status.

‘Other failing grade’ distribution and examination performance

The data for this study is from semester 1 2008 and semester 1 2009. The data that is accorded a prominent place in the analysis is that which concerns: (1) course abandonment rates as measured by ‘other fails’; (2) progression rates (one minus the percentage of ‘total fails’); and (3) examination performance. A fail grade is given for genuine low performance when a student has submitted all assessment items and sat the exam but scored less than 50% of the total available marks. However, ‘other’ failing grades for non-completion, not sitting the exam or no participation are given for incompleted studies. Course abandonment may be measured by the percentage of ‘other fails’. Because of the complex interplay of variables including students’ personal situations, reducing the number of ‘other fails’ is among the most difficult tasks for the university. “Significant differences in ‘other fails’ across semesters are likely to indicate substantial improvements or, conversely, problems with these facets of a course” (Phillips and Loch, 2009).

Table 1: Other Fails 2008 and 2009; from (Phillips and Loch, 2010).

<table>
<thead>
<tr>
<th></th>
<th>2008 Total enrolments</th>
<th>Other Fails (% of Students)</th>
<th>2009 Total enrolments</th>
<th>Other Fails (% of Students)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toowoomba (On Campus) Students</td>
<td>89</td>
<td>15.85%</td>
<td>99</td>
<td>6.12%</td>
</tr>
<tr>
<td>Fraser Coast (On Campus) Students</td>
<td>12</td>
<td>0.00%</td>
<td>13</td>
<td>0.00%</td>
</tr>
<tr>
<td>Springfield (On Campus) Students</td>
<td>24</td>
<td>40.91%</td>
<td>40</td>
<td>20.00%</td>
</tr>
<tr>
<td>Distance (External) Students</td>
<td>258</td>
<td>15.60%</td>
<td>254</td>
<td>12.70%</td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>16.67%</td>
<td>406</td>
<td>11.42%</td>
</tr>
</tbody>
</table>

Care needs to be taken when different cohorts are compared. However, the difference in ‘other fails’ across 2008 and 2009 is striking and cannot be entirely attributed to differences in the core characteristics of the two student cohorts. Final examinations are very similar in structure and difficulty and markers remained mostly unchanged across the two semesters (with the course instructor moderating the consistency of markers). The average examination performance is reported in Table 2.

Table 2: Examination Performance 2008 and 2009, from (Phillips and Loch, 2010)

<table>
<thead>
<tr>
<th></th>
<th>2008 Total enrolments</th>
<th>Examination Performance (Max. 100%)</th>
<th>2009 Total enrolments</th>
<th>Examination Performance (Max. 100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toowoomba (On Campus) Students</td>
<td>89</td>
<td>38.35%</td>
<td>99</td>
<td>66.94%</td>
</tr>
<tr>
<td>Fraser Coast (On Campus) Students</td>
<td>12</td>
<td>57.71%</td>
<td>13</td>
<td>75.15%</td>
</tr>
<tr>
<td>Springfield (On Campus) Students</td>
<td>24</td>
<td>30.72%</td>
<td>40</td>
<td>58.17%</td>
</tr>
<tr>
<td>Distance (External) Students</td>
<td>258</td>
<td>35.58%</td>
<td>254</td>
<td>64.11%</td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>36.61%</td>
<td>406</td>
<td>64.56%</td>
</tr>
</tbody>
</table>
Socio-economic status

Attention may now be turned towards the performance and course abandonment statistics that characterise particular SES groups within the class. The classification of students in low, medium and high SES is undertaken using information provided by the Department of Education, Employment and Workplace Relations (DEEWR) and the Australian Bureau of Statistics (ABS). In the ABS Socio-Economic Indexes for Areas (SEIFA) (ABS, 2008), the Index of Education and Occupation (IEO) ranking of postcodes is of relevance to this study. The postcodes that comprise the bottom 25 percent of median household incomes of the population aged between 15 to 64 years, are considered low SES postcodes. The University receives a postcode/SES status assignment (low, medium, high, other (overseas)) for each student from the Department of Education, Employment and Workplace Relations (DEEWR) via the ABS, based on the postcode of a student’s home residence. It is the Department’s classifications that are, therefore, deployed in this analysis. It should be noted that the use of postcodes to establish SES is currently under review by the Australian government.

In 2008, there were 383 students enrolled in the introductory finance course at USQ. In 2009, there were 406 students enrolled. ‘Partner’ enrolments (where students are enrolled through an educational partner institution and are provided with ‘local’ tutoring) will be excluded in the following analysis, and without those there were 285 students enrolled in semester 1 2008 and 311 enrolled in semester 1 2009. More than one-third of these students are classified as low SES in 2009, and just under a third in 2008. The SES classification for all students in 2008 and 2009 is reported in Table 3:

<table>
<thead>
<tr>
<th>Socio-economic Status</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>38</td>
<td>36</td>
</tr>
<tr>
<td>Medium</td>
<td>140</td>
<td>155</td>
</tr>
<tr>
<td>Low</td>
<td>92</td>
<td>104</td>
</tr>
<tr>
<td>Other [no info - overseas]</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>285</td>
<td>311</td>
</tr>
</tbody>
</table>

In order to provide as much information as possible about the characteristics of the student cohorts, it is worthwhile examining the students tertiary entrance (high school) scores (OPs or overall positions) and the distribution of OP scores across SES (OP Score is calculated as a value between 1 and 25, 1 is the highest score). Because many students at the University of Southern Queensland are mature age or enter via ‘alternative pathways’, a large percentage of students do not have an overall position. This information is presented in Table 4. Note that the percentages in the far right-hand columns are based on the number of students in each SES category. There are (from Table 3), for example, 155 students with medium SES backgrounds in 2009 out of which 38 received OP 11 to 15. That is 25 percent of the total number of students in that SES category. The lower OP brackets are highlighted for each of the SES backgrounds in Table 4.

<table>
<thead>
<tr>
<th>Socio-economic Status</th>
<th>2008 ( % by SES Category)</th>
<th>2009 ( % by SES Category)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3 8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>6 16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>10 26</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>1 3</td>
<td>6</td>
</tr>
<tr>
<td>No OP</td>
<td>18 47</td>
<td>33</td>
</tr>
<tr>
<td>Medium</td>
<td>24 17</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>39 28</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>34 24</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>7 5</td>
<td>3</td>
</tr>
<tr>
<td>No OP</td>
<td>36 26</td>
<td>25</td>
</tr>
<tr>
<td>Low</td>
<td>12 13</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>27 29</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>28 30</td>
<td>29</td>
</tr>
</tbody>
</table>
Previous studies provide ample reason for us to expect that students from socio-economically disadvantaged backgrounds will exhibit lower levels of academic achievement. The literature also provides reasons to expect that any relationship between (pre-university) academic achievement and SES is unlikely to be completely clear-cut. The data presented in Table 4 are in accordance with these expectations. Students with low SES backgrounds exhibit a much higher percentage of OP scores in the 11 to 19 brackets. Indeed, in 2009, 47 percent of students from low SES backgrounds have OP scores in these lower categories. This compares to 28 percent and 34 percent for medium and high SES background students respectively. The students from low SES backgrounds enrolled in the introductory corporate finance course have, generally, exhibited lower academic achievement prior to entry into higher education.

As can be expected, in light of the course abandonment and exam performance statistics already presented, the performance of students overall improved during 2009. To formally compare the non-tablet semester 1 2008 with the tablet semester 1 2009 in the context of SES, we focus on the progression rates for each SES group. The progression rate is one minus the percentage of total failing grades for the course. If ‘total’ fails are 60 percent, for example, the progression rate for the course will be 0.40. A higher progression rate indicates that students have both stayed with the course (no ‘other fail’ grade) and achieved a passing grade or higher. The progression rates by SES group are reported for semester 1 2008 and semester 1 2009 in Table 5. The progression rates by SES group and OP score for semester 1 2008 and semester 1 2009 are reported in Table 6.

<table>
<thead>
<tr>
<th>OP 16 – 19</th>
<th>10</th>
<th>8</th>
<th>11</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>No OP</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>5</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>OP 6 – 10</td>
<td>4</td>
<td>4</td>
<td>27</td>
<td>25</td>
</tr>
<tr>
<td>OP 11 – 15</td>
<td>5</td>
<td>2</td>
<td>33</td>
<td>13</td>
</tr>
<tr>
<td>No OP</td>
<td>3</td>
<td>5</td>
<td>20</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 5: Progression Statistics by SES

<table>
<thead>
<tr>
<th>Socio-economic Status</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.63</td>
<td>0.58</td>
</tr>
<tr>
<td>Medium</td>
<td>0.69</td>
<td>0.83</td>
</tr>
<tr>
<td>Low</td>
<td>0.68</td>
<td>0.90</td>
</tr>
<tr>
<td>Other [no info - overseas]</td>
<td>0.53</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 6: Progression Statistics by SES and OP Score

<table>
<thead>
<tr>
<th>Socio-economic Status</th>
<th>OP group</th>
<th>Progression</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>OP 1 – 5</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>OP 6 – 10</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>OP 11 – 15</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>OP 16 – 19</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>No OP</td>
<td>0.67</td>
</tr>
<tr>
<td>Medium</td>
<td>OP 1 – 5</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>OP 6 – 10</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>OP 11 – 15</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>OP 16 – 19</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>No OP</td>
<td>0.69</td>
</tr>
<tr>
<td>Low</td>
<td>OP 1 – 5</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>OP 6 – 10</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>OP 11 – 15</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>OP 16 – 19</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>No OP</td>
<td>0.73</td>
</tr>
<tr>
<td>Other</td>
<td>OP 1 – 5</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>OP 6 – 10</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>OP 11 – 15</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>No OP</td>
<td>0.33</td>
</tr>
</tbody>
</table>
The statistics reported in Tables 5 and 6 are equally as striking as those regarding course abandonment and examination performance. Several features in Table 5 bear further comment. First, there is a substantial improvement between 2008 and 2009 among both medium and low SES background students. Second, the largest improvement is exhibited by the low SES students. Third, high SES students actually exhibited a slight decline (although probably not statistically significant and the sample size for high SES is small enough to be susceptible to cohort idiosyncrasies). Among the students classified as low or medium SES, there was an approximately 42 percent improvement in course progression when semester 1 2008 and semester 1 2009 are compared (weighted to adjust for increases in student numbers in 2009). Fourth, the largest improvements are in those lower OP-score categories within low and medium SES categories. Students exhibiting lower academic achievement prior to entering university and hailing from low and medium SES backgrounds have exhibited a much higher level of progression during the Tablet PC semester. The Tablet PC semester 1 2009 is characterised by a very substantial improvement in overall course progression. This is a very encouraging set of results.

Discussion and future research

Following the new approach to the delivery of the introductory corporate finance course, performance and engagement of students with low and medium SES backgrounds improved substantially, and more of these students stayed in the course and successfully completed it. The most important issue to discuss in this section is the obstacles that may have been offset or overcome by the utilisation of the technology and the lecture delivery approach that it facilitated. This question cannot be answered completely at this stage but its discussion sow the seeds for future research. Of particular interest is the way in which teaching technologies may interact with self-efficacy for self-regulated learning. Self-efficacy for self-regulated learning has received a great deal of attention within the literature. As pointed out earlier, self-regulatory efficacy may be influenced by SES factors and lower self-regulatory efficacy negatively impacts student performance. Within the general student population at USQ and, as highlighted in the previous section, within the particular group of students enrolled in the introductory corporate finance course, students from low SES backgrounds constitute a considerable portion of the student body. Many of these are enrolled for distance education and must balance the pursuit of higher education with various additional challenges (many of which, no doubt, are related to SES context).

In this environment, self-efficacy for self-regulated learning would seem to be of some considerable importance. There are several reasons why self-regulatory efficacy may be relevant to this discussion:

- First, studying relatively independently by distance education is likely to test students’ self-efficacy and self-regulation. Caprara et al. (2008, p.532) highlight the importance of preventing ‘erosion’ in self-efficacy and self-regulatory efficacy: “[A] decline in self-regulatory efficacy foreshadows low academic performance and school drop out.” Using technology to reach students studying in distance mode may be critical in overcoming the erosion of self-regulatory efficacy.

- Second, as Caprara et al. (2008, p.525) highlight, “[…self-regulatory efficacy has] ‘growing primacy in contemporary life.’ Information technologies are globalising knowledge and altering educational systems…. In this new era, the construction of knowledge will rely increasingly on electronic inquiry.” In an environment where technology is used to deliver higher education, determining how to use technologies effectively and in a manner that recognises this primacy, allows students to build their self-regulatory efficacy and broadens students’ horizons beyond the limits set by SES, and social and peer groups may be of significant importance.

- Third, Caprara et al. (2008, p.527) point out that SES may influence academic achievement directly and indirectly through its impact on self regulatory efficacy. What might be considered thoroughly in future research, therefore, is whether the positive effects that have been observed in the present trial of Tablet PC technology are in some way due to the overcoming or offsetting to some degree of negative SES (and other) influences on self-regulated efficacy by the effective utilisation of teaching technology.

The efficacy of teaching technologies in improving student performance is still an open research program. Indeed, so is the study of technology, SES and academic achievement. There are numerous challenges facing researchers who seek to unravel the complex interaction and cumulative effects of many different variables. The extant research is small but enough information has been gathered to suggest that institutions and instructors attempting to overcome or offset the negative influence of SES may find some assistance in the appropriate and effective deployment of teaching technology. For example, Joo, Bond and Choi (2000) and Debowski, Wood and Bandura (2001) found that established relationships between self-efficacy and academic achievement apply with similar force in web-based
learning contexts. It is possible that at least some portion of the positive outcomes generated for low and medium SES students during the Tablet PC trial described in this paper might be accounted for by considering the offsetting effects of the technology on the negative influences of SES on self-regulatory efficacy. The opportunity to experience elements (audio and video) of a positive classroom experience may offset (or help to overcome) the negative influences that other factors (including those stemming from SES context) may have on self-regulatory efficacy and academic achievement. It is in this way, perhaps, that the effective utilisation of teaching technology can help to improve student motivation, engagement and performance. Deeper inquiry into these issues is a task for future research.

Conclusions

This paper presents the results of a Tablet PC trial within an introductory corporate finance course at USQ. The Tablet PC facilitated the interactive construction of lectures in class and the distribution of screencasts to all students via the LMS. Of particular interest, both in light of Australian Government higher education policy and the particular demographics that characterise our institution, is the way in which teaching technology can be used to enhance the motivation, engagement and performance of students from low and medium SES backgrounds. The comparison of course abandonment statistics, examination performance and progression statistics point to a considerable improvement between the non-tablet semester 1 2008 and the tablet semester 1 2009. At least some of this improvement must be attributable to the change in delivery approach during 2009. Importantly, students from low SES backgrounds exhibited the greatest improvement in progression (staying with the course and obtaining a passing grade or better) in 2009. This result was mirrored by the medium SES students while the high SES students’ progression remained fairly static. It is possible that the effective utilisation of teaching technologies helped to offset some of the challenges to self-regulatory efficacy (which is important for distance study) presented by SES and other factors. This is a promising prospect for future research.

References


Generic Educational Knowledge Representation for Adaptive and Cognitive Systems

Arturo Caravantes and Ramón Galán

Institute of Educational Sciences, Universidad Politécnica de Madrid, Madrid, Spain // 1Intelligent Control Group, Universidad Politécnica de Madrid, Madrid, Spain // arturo.caravantes@upm.es // ramon.galan@upm.es

ABSTRACT

The interoperability of educational systems, encouraged by the development of specifications, standards and tools related to the Semantic Web is limited to the exchange of information in domain and student models. High system interoperability requires that a common framework be defined that represents the functional essence of educational systems. To address this need, we developed a generic model of educational systems that we called Cognitive Ontology of Educational Systems (COES) and we encoded it as functional reference ontology. It generalizes the educational system architecture, taking into account the cognitive perspective. This framework ranges from the usual e-learning systems to complex cognitive and adaptive hypermedia systems. This article describes the part of the COES related to domain knowledge representation and proposes an implementation called Generic Educational Knowledge (GEK). The GEK model is structured in a flexible way that allows the authors to codify instructional and semantic levels as needed. In order to test its feasibility, the model was applied in a distance learning course using two educational-knowledge models, GEK and ADL-SCORM, and two educational systems, a traditional Web-Based Adaptive Hypermedia System and a Rich Internet Application with dialog interaction and cognitive monitoring.

Keywords

Intelligent educational systems, Cognitive systems, Knowledge representation, Domain model

Introduction

The development of specifications, standards and tools related to the Semantic Web (Hendler, 2001; Dolog & Nejdl, 2007) has led to the development of ontologies with the objective of globally interconnecting knowledge. Applying this knowledge to education requires the development of browser technologies that automatically filter knowledge in order to generate an appropriate offer for each educational objective. Existing e-learning management systems (LMS) use instructional knowledge coded with different levels of granularity and associated with multimedia resources such as ADL-SCORM or IMS-CC. However, research and development centers propose adaptive models that range from basic instructional rules to implicit and complex representations related to the needs of each particular implementation. Currently, efforts are focused on the creation of models, specifications and standards that increase computational integration (Kozaki, Hayashi, Sasajima, Tarumi, & Mizoguchi, 2008; Dietze, Gugliotta, & Domingue, 2007; De Roure & Hendler, 2004) using ontologies (Gaeta, Orciuoli, & Ritrovato, 2009; Žitko, Stankov, Rosić, & Grubišić, 2009; Isotani & Mizoguchi, 2008; Boyce & Pahl, 2007; W3C, 2004) in what is called the Intelligent/Semantic Web (Devedzic, 2004; Zhong, Liu, & Yao, 2002) or Web3.0 (Zeldman, 2006).

This approach leads to the emerging paradigm of open-corpus knowledge (Sosnovsky, 2009). It offers a wealth of knowledge described at a semantic level to educational systems, which greatly improves the capabilities of automatic systems. However, the interoperability of learning platforms is restricted to the exchange of information in domain and student models that follow specifications. We believe that the interoperability of educational systems on the Grid requires a framework to facilitate interaction at a functional level and to overcome the limitations of the specifications. The LTI project (IMS, 2009) is an example of such efforts.

Work on Intelligent Educational Systems (IESs) is traditionally divided into two main paradigms (Nicholas & Martin, 2008): Intelligent Tutoring Systems (ITSs) (Murray, 1999; Wenger, 1987) and Adaptive Hypermedia Systems (AHSs) (Dolog, 2008; Brusilovsky, 1996). The former are designed to guide students in acquiring specific skills to complement the primary instructional method and use Constraint-Based Modeling from Ohlsson’s theory. The system works by correcting the learner’s practice errors and guiding him/her to improve; the WETAS system is an example (Martin, Mitrovic, & Suraweera, 2008). AHSs, in contrast, are directed to a complete theoretical instruction through learning concepts. They use intelligent systems to adapt the sequence, presentation and contents according to the student model. New educational systems incorporate the latest advances in cognitive science. Some of them employ estimated cognitive parameters of learning, while others develop computing architectures that...
simulate the mental processing of students. Cognitive control is the ability to integrate information from a multitude of sources and use that information to flexibly guide behavior. From the characteristics described above, we developed a framework for educational systems called the Cognitive Ontology of Educational Systems (COES) whose main objective is to provide functional interoperability. The teaching-learning process is a complete and indivisible entity in which numerous real and abstract elements are artificially bound together by functions in different areas: emotional, cognitive, instructional, behavioral, etc. The traditional IES architecture (De Bra, Houben, & Wu, 1999; Murray, 1998) has kept a functional division in which a processor uses the rules of a pedagogical/adaptive model to select content from a domain model. However, many systems encode pedagogical models through specific operational rules within the domain model. In the COES proposal, we group the two models into one domain called the \textit{educational domain}. Within the architecture of the three domains shown in Figure 1, the \textit{interaction domain} facilitates communication with the user and agglutinates the user interface and the control unit. The \textit{educational domain} encodes knowledge involved in the teaching-learning process; and the \textit{personal domain} estimates the characteristics and state of the student and simulates his behavior.

![Figure 1: Domain division in COES](image)

In section 2, we describe the part of the COES model related to the educational domain and propose a framework of the educational domain that is compatible with traditional models and that incorporates the cognitive point of view. In section 3, we analyze the different ways of encoding educational knowledge and propose a model called Generic Educational Knowledge (GEK), which is compatible with the COES framework. Section 4 describes the results of a distance learning course supported by two COES-compatible educational systems that use two ways of representing the educational knowledge: GEK and ADL-SCORM.

**Educational Domain**

The essential part of the teaching-learning model is the educational domain, which is responsible for the stimuli generated toward the student to properly modify his cognitive state. All intelligent educational systems incorporate an educational domain that determines its scope and the process effectiveness. Usually the educational domain is divided into two parts: one represents the target knowledge of the learning process and is called domain model, and the other defines rules or procedures governing the process and is called the pedagogical/adaptive/operational model. We elaborated on an ontology named the COES which aims to homogenize and generalize the architecture of IES into the cognitive point of view, extending models like AHAM (De Bra, Houben, & Wu, 1999), Munich (Koch & Wirsing, 2002), GAF (Knutov, 2008), OMNIBUS (Hayashi, Bourdeau, & Mizoguchi, 2008) or ODAS (Tran, Wang, Lamperter, & Cimiano, 2008). In this regard, we have grouped these modules into one, because they are very strongly related, and in many cases, the representation is a single one. Figure 2 shows the class diagram of the COES educational domain.

The main part of the educational domain is the internal representation of educational knowledge. This knowledge is the perceptual-semantic device that teachers use to influence student cognitive states. Therefore, it will always be a structure of media resources linked in some way with a conceptual meta-structure. In order for the system to have the versatility that the teaching-learning process requires, the educational knowledge representation should have some of the following characteristics: categorized, structured, coherent, continuous, adaptive, regulated, reusable and interoperable.

The teaching process is limited to the presentation of adapted sensory stimuli to the student. These stimuli are dynamically generated or selected from a set of \textit{Educational Resources} (\textit{ER}s). An \textit{ER} is composed of a sequence of instructional actions with implicit learning processes involved. These vary from the presentation of text, image and sound stimuli to simulators, chats and shared applications. The selection procedure of \textit{ER}s is based on metadata parameters, especially on pedagogical features that are specified in the resource type (channel, audience, strategy…).
Following the above recommendations and the educational models analyzed, in the COES we suggest a traditional knowledge representation based on a network of Educational Knowledge Nodes (EKNs). They are structured in a network of nested elements using Educational Knowledge Relations (EKRs) that may point to other nodes (EKNs) or resources (ERs). The network generates a meta-structure of ERs that systems display in the user interface to act on the learner’s cognitive state. The structure’s granularity and complexity level depend on the degree of inclusion of some nodes into others.

The proposed COES educational knowledge spans from simple resources associations to more complex representations in a semantic network. It is more generic than common specifications like ADL-SCORM and IMS-CC but is compatible with them. The main difference is that an EKN makes it possible to establish relationships between all levels: didactic, instructional, conceptual, competence, and so on (see next section). This way, designers can codify educational knowledge as symbolic as they consider making knowledge structure more stable and letting it remain unchanged in each instructional session. We could say that traditional models allow the representation of domain contents, while the COES model facilitates the encoding of domain knowledge.

In order to update and keep track of the student’s knowledge you need to establish a link between the educational and cognitive domains. In the COES model, this relationship is made by Cognitive Sentences (CSs). These are descriptions of the cognitive state that are processed when the student plays an educational knowledge node (EKN). In other cases, these sentences work as prerequisites for the activation of an educational node. Cognitive sentences are an implementation form of Insights (Ausubel, Novak, & Hanesian, 1983), Chunks (Anderson & Lebiere, 1998) or Propositions (Pozo, 2003). They make cognitive changes and, in doing so, should manage certainty measures.

Teachers use the educational knowledge of a particular domain together with a meta-knowledge that encodes pedagogical skills for process controlling. This pedagogical knowledge represents instructional principles such as positive reinforcement, variability, action, etc., that facilitate the proper selection and sequencing of contents. It is a type of procedural knowledge based on Pedagogical Regulators (PRs) that controls the process, reading information from the emotional, characteristic and instructional domains. Common IESs implement an explicit pedagogical knowledge just as expert systems do using PRs called logical rules. However, there are other valid implicit representations such as neural networks, in which the PRs are called nodes, artificial neurons, and so on.

The control of the educational domain is performed by the Educational Processor (EP), which receive actions from the interaction domain and generate a response by consulting the emotional, characteristic and cognitive states. The educational domain communicates with the interaction domain through a language based on elements called Educational Actions (EAs): show content, reply, consult, advise, motivate, evaluate, and so on.

The operation of the educational processor requires the maintenance of a working memory that records the educational strategies and techniques being used and the states of active educational resources (ERs) and nodes (EKNs). The processor consists of three memory subsystems: Educational Domain Stable Memory (eDSM) a long-term declarative type that records the domain knowledge; Educational Pedagogical Stable Memory (ePSM) a long-
term procedural type that works with process regulators; and *Educational Temporary Memory (ETM)* a short-term declarative type that keeps process states. An *Educational State (ES)* collects all information about the activation of a node (KN) for a student during an instructional session. The state is the sum of *Educational Features (EFs)* such as: repeated, read, viewed, known, evaluated, learned, strategy used, etc.

**Generic Educational Knowledge**

The educational domain handles transmitter-knowledge used by a teacher in his activity, whereas the cognitive domain tries to represent transferred-knowledge in the learner model. Usually, many ways exist to explain the same concept; therefore educational knowledge is an explicit temporal representation of the content created by a teacher to deliver a piece of knowledge at the cognitive level. It could be said that the cognitive domain deals with Thinking-oriented Knowledge (ToK) while the educational domain works with Learning-oriented Knowledge (LoK). Knowledge representation uses conscious/explicit/symbolic elements that facilitate logical reasoning and unconscious/implicit/subsymbolic elements related to automated processing. This division should also be implemented in the educational and cognitive domains through several levels as Figure 3 shows. Educational knowledge is codified in didactic elements that contain instructional resources and activities. It is explicit and works sequentially. Knowledge in the cognitive domain depends on the features of the internal processors in the human mind and it works in parallel. It is associative-implicit and codified in conceptual elements based on perceptive parameters.

![Figure 3: Knowledge comparison: ToK - LoK](image)

In this work, we put forward a flexible educational knowledge representation, directly linked to the cognitive domain, which facilitates student tutoring and monitoring. It is stored in a knowledge base, and it is encoded at different levels that determine the control possibilities of the teaching and learning. The level of representation establishes the features of the control variable and affects the pedagogical/adaptive capacity of the system. Figure 4 summarizes the different views of the educational knowledge representation from the physical level to the logical one (Chi, 2009; Benyon, 1993), and their connections with learner monitoring.

The resource layer is the basic representation needed for educational system operation. The granularity scale and the encoding format of resources are the main factors that determine the adaptation possibilities of the presentation. Many systems incorporate a level of specification called didactic/instructional to extend the adaptability to the sequence. This layer defines the methodology of the instructional process through a hierarchy of contents and activities with structural relationships, as in RLATES/IGNATES (Iglesias, Martínez, Aler, & Fernández, 2008) or ZOSMAT (Keles, Ocak, Keles, & Gülcü, 2009). This level of specification usually incorporates prerequisites rules that determine the instructional flow (Gagné, Briggs, & Wagner, 1988) as in SmartTutor (Cheung, Hui, Zhang, & Yiu, 2003). Furthermore, the resource adaptability is achieved through flexible links between instructional elements and the assets that map them.
Because there are many instructional ways of presenting, structuring and teaching the same concept, systems establish a higher level of representation to unify student tracking, which we call the competence layer, where learner knowledge and the skills to achieve are specified. At this level of representation, instructional elements known as LO (Learning Objects) exist, such as IEEE-LTSC, ADL-SCORM (ADL, 2004), IMS-LD (2003), IMS-CC (2008), RLO (CISCO Systems, 2003), ePack-Blackboard and specifications as LMML (Süss, 2001), LAMS (LAMS Foundation, 2002) adopted by Moodle, PALO (Rodriguez-Artacho & Verdejo, 2004) and the ALOCoM model (Verbert, Duval, Meire, Jovanovic, & Gasevic, 2006). This object package is a set of structured resources, metadata described, with an implicit methodology and learning objectives. Such objects improve the learning process by distributing and sharing instructional resources. However, they are not being used as expected, probably because of the lack of consensus on a common core ontology (Balatsoukas, Morris, & O’Brien, 2008) and the viability of many proposals. The LAOS model (Cristea & De Mooij, 2003) is an example of this level of representation.

Competencies and objectives are artificial taxonomies related but not equivalent to the real knowledge. To improve monitoring of the student’s cognitive skills, authors establish a new layer where knowledge is represented at the conceptual level. Samples of this layer can be found in the AIMS system (Aroyo, Mizoguchi, & Tzolov, 2003; Aroyo, Denaux, Dimitrova, & Pye, 2006) and the TEx-Sys system (Stankov, Rosic, Zitko, & Grubisic, 2008).

Ontologies and concept maps, disciplines that facilitate knowledge representation at the semantic level, have become ones of the development lines of educational knowledge encoding in recent years (Chi, 2009; Martin, Mitrovic, & Suraeweera, 2008; Boyce & Pahl, 2007; Coffey, 2007; Devedzic, 2006), especially in application to the Semantic Web (Berners-Lee, Hendler, & Lassila, 2001). This conceptual-semantic way of representing the domain model are used in the proposal of (Papasalouros, Retalis, & Papaspyrou, 2004), the Inca architecture (Nabeth, Razmerita, Angehrn, & Roda, 2005), the CAM (Hendrix, De Bra, Pechenizkiy, Smits, & Cristea, 2008) and Hera (Houben, et al., 2008) models, and the IWT (Gaeta, Orciuoli, & Ritrovato, 2009) and ICDS (Lee, Lee, & Leu, 2009) systems.

The representation of educational knowledge directly determines the capacity of IESs. Basic systems represent knowledge in a single asset that implicitly encodes the didactic, methodological and cognitive structure. Monitoring, if it exists, is limited to logging the inputs and resource access times. Advanced systems use cognitive recording (what has been learned) and instructional tracking (what has been done). In those systems, knowledge is encoded in a highly fragmented way, interlinked and metadata described. Knowledge components are organized into complex didactic, competence and conceptual structures so as to improve the monitoring and allow the adaptive process of dynamic content generation; they clearly distinguish resources from their organization and use.
The more representative educational systems share a common domain model structure with a number of specification levels. Depending on the needs of the system (Chi, 2009; Sosnovsky, 2006; Aroyo, Mizoguchi, & Tzolov, 2003) we find the following: the physical/perceptual layer of resources, the instructional/didactic layer, the methodological layer and the semantic layer. Sometimes, several optional intermediate levels of representation are inserted to facilitate system processing, as shown in Figure 5.

Following the suggested COES and the structural scheme (Figure 5), we propose a Generic Educational Knowledge (GEK) mapped into an eKN/eKR network as show in Figure 6. It is organized from the top abstract elements called the nucleus to the bottom specific resources. Every relation can point to a specific node or can include a selector script that allows searching for the appropriate node from databases or remote repositories. In addition, relations may incorporate a filter to select parts of the referenced node.

Learning content specifications like ADL-SCORM (ADL, 2004) or IMS-CC (2008) are based on independent and interchangeable objects or packages that encode one or more ways to teach-learn something. However, we have opted for a closer representation to the cognitive domain, so that when the system detects a deficiency in the student cognitive state, it may resort to the educational domain. Furthermore, each educational node (eKN) is not a complete and independent object that can be packaged and shared with all its functionality. Frequently, following the open corpus trend, educational nodes include flexible relationships that refer to items with certain parameters that must be sought outside the node coding itself. This ensures that knowledge is organized with a higher semantic charge, adaptive capacity and component reusability.

At the highest semantic level of the educational knowledge structure is a collection of nodes called the SEN (Semantics and Epistemological Nucleus). These elements can be conceptual nodes or didactic-structural elements that teachers create to facilitate the learning process. They are identified by a globally unique identifier and metadata, which are used to construct flexible links. The metadata includes information about type (conceptual/didactic), title, abstraction level and keywords. A SEN is equivalent to the class/entity of different modeling forms of didactics and semantics. SEN content is defined by Semantic Relations with other SENs. They are used to establish subordination, generalization and combination with other SENs to express WhatIs, WhatHas, WhatHappens, HowWorks, HowIsMade, and so on. SENs are complemented with both teaching (how to explain) and perceptual (how to show) information. This information is linked through Instructional Relations that establish the function (Guidance/Introduction/Content/Documentation/Assessment) of a node in the lower levels.
The methodological layer is composed of methods and activities packed in PROCESS nodes. Each of the nodes encodes a specific way of teaching/learning the meaning of its parent SEN, and so they resolve the need for different ways to teach/learn the same concept to extend the acting capabilities in the scope of multiple intelligences (Gardner, 1993) and to enable the SEN adaptive processes to the learner. For example, authors could plan a learning process that follows Bloom’s levels progressively through the selection of the corresponding PROCESS (Hwang, Wang, Hwang, Huang, & Huang, 2008). The definition of PROCESS nodes could include prerequisites and metadata that allow database searching through type, instructional method, student characteristics, teaching strategies and keywords. They are encoded as a flow of ACTIVITY nodes that controls the presentation sequence of the environments that interact with the student. PROCESS activation runs in a hierarchical space through Action Relations coded in terms of predecessor-successor. Depending on the activity type, the interaction can occur from the resources to the student (presentations), between the student and the resources (documents), between the student and his classmates (discussion) and between the student and the system (tutoring). When an ACTIVITY node requires the use of assets, it includes Content Relations with nodes in the lower layers.

The content layer consists of didactic structures (Documents/Presentations/Questionnaires...) defined by ORGANIZATION nodes, which are equivalent to Organization elements in ADL-SCORM or Environment nodes in IMS-LD. Each ORGANIZATION contains instructional elements called STIMULI that interact with the student to achieve an educational objective. Sequence Relations define the clusters, the order of presentation, the activation conditions and the estimated duration of the referred stimuli. A stimulus is a set of physical resources (text, image and sound) that follow educational goals to change the student’s cognitive state. They are equivalent to the Item elements of ADL-SCORM and can incorporate tracking options such as SCOs (Sharable Content Objects) or simply be presented as an asset. Cognitive and aptitude monitoring is implemented through Cognitive Sentences (CSs) that trigger when a student successfully completes an ACTIVITY or a STIMULUS. They can record new knowledge to the student's cognitive domain or change the certainty of existing knowledge.

Distance learning has usually made available to the student a set of continuous learning resources (documents, presentations, videos, animations and simulations) and discrete resources (texts, graphics, links and images) (Verbert & Duval, 2004). Resources are located at the lower material level of the educational knowledge structure to interact with the student to transfer knowledge. Educational resources are increasingly specified by soft links (Brusilovsky & Henze, 2007) that allow them to be searched, filtered and selected from large and dynamic repositories using metadata such as type (exercise, questionnaire, diagram, graph, table, text...), format, language, difficulty, etc. (see IEEE-LOM).

Educational knowledge elements (except resources) encode data, but not its interface instantiation. Any node that can be shown on the user interface may include Presentation Relations with resources, such as backgrounds, icons, titles, covers, etc., to make it more intuitive and adapted to the learner's style.
The creation of a GEK base usually begins with a source document located in the bottom psychical/material layer. This is broken into subdocuments that are structured by creating nodes in the upper levels. In some cases the process is reversed; for example, authors can begin with an ontology that is supplemented with new relations to the nodes in the lower layers.

Educational knowledge codification is achieved by any method that makes it possible to incorporate some elements into others. Implementation through a relational database improves query and update performance, compared to XML-based languages like RDF/OWL. We opted for a database for local educational nodes and an XML coding for the nodes in remote repositories (Figure 7). In any case, we can move from one encoding to another depending on the system needs.

Evaluation

The GEK model was evaluated by an e-learning course in 2008 and 2009 about Web Design in the Teacher Training area. The course was worth one European Credit (ECTS) and held at the Institute of Educational Sciences in the...
Technical University of Madrid. To check the interoperability and effectiveness of the GEK model (Figure 8), we used two educational systems: an adaptive educational hypermedia system named TIX (eXtensible Intelligent Tutor) (Bravo & Caravantes, 2004) and a cognitive educational system called MAP (Adaptive Pedagogical Module). Both systems consult learning elements stored in a GEK repository about Web Design. As a reference model we used the same curriculum on Web Design encoded in SCORM1.2 LO hosted on Moodle. A total of 102 students applied for and attended the course in successive phases with different systems. However, only data from 87 students were used for the evaluation process because 15 students did not develop the commitment and continuity required. Of the 87 students in the evaluation, 17 students attended the course in Moodle, 25 in MAP, 29 in TIX and 16 in both MAP and TIX.

![Figure 8: Evaluation scheme: educational knowledge and systems](image)

TIX is a web-based system that controls the teaching-learning process through the activation of links in the user interface. It consists of an http-server application that uses the GEK base to generate HTML pages with links adapted to the student (Figure 9). It combines the specific control instructions introduced by a human tutor during the learning process with the automatic control programmed into implicit pedagogical rules.

![Figure 9: System architectures](image)

MAP is a Cognitive Tutor (Anderson, Corbett, & Koedinger, 1995) that aims to simulate the internal working of the student learning process, and thus increase the regulation scope that traditional AHS monitoring allows (Roda & Nabeth, 2007). The system consists of two components, as shown in Figure 9: a web client that performs the slave interface function as an RIA (Rich Internet Application) and a master server in charge of the processing of the student interaction and tracking. Essential control and processing are performed in the server, whose architecture is based on simulator elements called mental processors. Each of them plays a functional simulation of a learner mental
system in which, as illustrated by cognitive neuroscience studies, there are many highly interconnected and distributed subsystems. When a MAP server starts, usually triggered by an external signal, the main mental processor is instantiated and becomes the MAPTutor. Each time the MAPTutor identifies a new person during the interaction process, it creates another linked mental processor. When a learner disconnects, the MAPTutor removes the associated mental processor and stores its status in long-term memories.

The GEK base about Web Design (Figure 10) consists of 316 SEN nodes. Of them, 38 are conceptual nodes used as prerequisites and structured into 28 clusters, and 241 are conceptual nodes of 66 clusters that encompass the course concepts. The remaining 37 are didactic nodes that shape the course. Overall, the course handles 242 cognitive sentences, among which 166 are assessed. From the home didactic node of the course, systems analyze semantic information (Jovanovic, et al., 2007) and build two semantic knowledge bases of both the didactic and the conceptual type (Figure 10), which make up the educational domain of systems. The GEK base, SCORM objects and item bank were made by a teacher-author engaged full time for 3 months.

Following the GEK structure (Figure 6), the 278 SEN nodes that shape course knowledge are described by relationships with nodes at lower levels. The load of levels, summarized in Figure 11, shows a knowledge base that is balanced in content and material layers and shortly methodology described. The 0.22 methodology index is equivalent to having two methodology choices in one of nine SEN nodes.

Figure 10: Sample of a GEK representation about Web Design

Figure 11: The analysis index of the GEK representation about Web Design. The indices are calculated by normalizing relations with respect to the number of SEN nodes (278)
When a course begins, systems use assessment type relations of the GEK base to generate an initial questionnaire. The student's knowledge base is initialized with cognitive sentences associated with correct answers on the questionnaire. The TIX and MAP systems compare the course and student knowledge bases to personalize instruction. They make an open learner model interface (Bull, Ahmad, Johnson, Johan, Mabbott, & Kerly, 2008) with didactic and conceptual nodes that include the students’ learning scores so as to enable the students to adapt their learning process. Each node offers content, information and assessment links. The achievement of learning activities triggers cognitive sentences that fill the students' knowledge base and update the certainty degrees of learning. We created two complementary mathematical parameters for managing certainty: perception certainty (Cp) and knowledge certainty (Ck). The Cp of a STIMULUS node is updated according to the time spent (t_s) by the student and the estimated time (T_s) associated with the stimulus. According to the tests carried out there is a wide variability in the time spent by students on the same stimulus, estimated as ±2/3. The time conversion graph to Cp parameter is shown in Figure 12. It shows that when a student spends 1/3T_s, a Cp=50% is established, while a t_s greater than 5/3T_s generates a Cp=100%. Moreover, the Cκ parameter combines Cp and the results of assessment activities. The Cp parameter is transferred to Cκ in the range of 0–75% as Figure 12 shows. In addition, each positive assessment does increase Cp in 2/3 of the remainder to 100%. Thus Cκ=75% represents the limit between the perceived state and the known state.

Students developed learning activities over 10–20 days. When all nodes of the student's knowledge base surpass Cκ=75%, the system reports "course completed" and generates a final questionnaire. The results of the final questionnaire as a measure of academic performance (Figure 13) show significant differences between students that used a cognitive system (MAP) and those who attended the course through the Moodle platform. The results registered by students in the adaptive hypermedia system (TIX) fall among the others.

As an additional comparative variable we calculated the Instructional Effort as the average of the perception certainty (Cp) of the content nodes (no assessment, no practice, no information ...) viewed on screen (Cp ≥ 50%) and weighted with respect to their T_s. The results show significant differences indicating that students using the cognitive system (MAP) spent more time attending resources on the user interface than students in the adaptive hypermedia system (TIX).

![Figure 12: Graphics for the certainty parameters.](image1)

![Figure 13: Box-and-whisker plots for the comparative performances of systems.](image2)
Conclusions

In working with computers, system behavior depends on two main factors: architecture and content. In this article, we combined both factors and proposed two frameworks that shape the educational domain from the functional and knowledge representation views.

We described the educational domain of a reference ontology for educational systems that we call the COES. In addition we propose a reference model for generic educational knowledge called GEK that implements the COES and is compatible with the main specifications for e-learning like ADL-SCORM and IMS-CC. It supports a highly flexible representation by means of relationships between the elements in different layers, in an attempt to reduce the gap between existing LMSs, AHSs and the Educational Semantic Web.

The GEK model is based on two main layers that are pedagogically autonomous: a top semantic layer and a bottom material layer. The model is supplemented with two intermediate layers that explicate the educational information implicitly encoded in resources. These intermediate layers extend the possibilities of adaptive educational systems in the same way that digital systems increase the capabilities of analog systems.

The top semantic layer enables us to have an educational knowledge representation closer to the knowledge used by a human teacher that dynamically argues and generates the appropriate instructional interaction and selects the resources that best fit the educative strategy. Furthermore, the model is flexible and updatable; authors can dynamically include new information in the nodes at the corresponding level.

How the GEK model is used by educational systems depends on their capabilities, from the simple adaptation of presentation to the complex tutoring teaching at a cognitive level. We implemented the GEK model in two educational systems: a web-based adaptive hypermedia system (TIX) that controls link activation and a cognitive tutoring system with dialog interaction (MAP). The comparative analysis of the results (Figure 13) showed a higher instructional effort in MAP learners that led to a 12.8% higher academic performance than that found among users of the Moodle benchmark system.

The results obtained in the test course confirm the feasibility of the model. Educational knowledge in the GEK format allows a wide range of adaptive possibilities, depending on the coding effort made by the authors. It can be used indistinctly by different systems like adaptive hypermedia or cognitive systems. On the other hand, it supports encoding flexible links to distributed nodes and resources, following the open-corpus trend. It can also be interchangeably used in a wide range of educational systems that encourage the interoperability of systems and students.

We are now augmenting the descriptive burden on the semantic and methodology layers in order to extend the adaptive capabilities and enable direct semantic instruction without the use of physical resources.

Acknowledgements

This project was developed with the support and resources of the Education Sciences Institute at the Universidad Politécnica de Madrid. We would like to thank Silvia Murga and Juan Luis Bravo for helpful discussion and comments.

References


