Educational Technology & Society
An International Journal

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

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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

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

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

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

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

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

Advisory board
Ignacio Aedo, Universidad Carlos II 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; Talc-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 Harlley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Rudd Leelouche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabama at Birmingham, USA.

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

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

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

Abstracting and Indexing

Guidelines for authors
Submissions are invited in the following categories:
• Peer reviewed publications: Full length articles (4000 - 7000 words)
• Book reviews
• Software reviews
• Website reviews
All peer review publications will be refereed in double-blind review process by at least two international reviewers with expertise in the relevant subject area. Book, Software and Website Reviews will not be reviewed, but the editors reserve the right to refuse or edit review.

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

Submission procedure
Authors, submitting articles for a particular special issue, should send their submissions directly to the appropriate Guest Editor. Guest Editors will advise the authors regarding submission procedure for the final version.
All submissions should be in electronic form. The editors will acknowledge the receipt of submission as soon as possible.
The preferred formats for submission are Word document and RTF, but editors will try their best for other formats too. For figures, GIF and JPEG (JPG) are the preferred formats. Authors must supply separate figures in one of these formats besides embedding in text.
Please provide following details with each submission: • Author(s) full name(s) including title(s), • Name of corresponding author, • Job title(s), • Organisation(s), • Full contact details of ALL authors including email address, postal address, telephone and fax numbers.
The submissions should be uploaded at http://www.ifets.info/ets_journal/upload.php. In case of difficulties, please contact kinshuk@ieee.org (Subject: Submission for Educational Technology & Society journal).
## Table of contents

### Special Issue Articles

- **Studying Research Collaboration Patterns via Co-authorship Analysis in the Field of TeL: The Case of Educational Technology & Society Journal**  
  *Panagiotis Zervas, Asimenia Tsitmidelli, Demetrios G. Sampson, Nian-Shing Chen and Kinshuk*  
  1–16

- **A Review of Research on Mobile Learning in Teacher Education**  
  *Evrim Baran*  
  17–32

- **Methodological Issues in Mobile Computer-Supported Collaborative Learning (mCSCL): What Methods, What to Measure and When to Measure?**  
  *Yanje Song*  
  33–48

- **Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence**  
  *Zacharoula Papamitsiou and Anastasios A. Economides*  
  49–64

- **Review of Speech-to-Text Recognition Technology for Enhancing Learning**  
  *Rustam Shadiev, Wu-Yuin Hwang, Nian-Shing Chen and Yueh-Min Huang*  
  65–84

- **The Influence of Young Children’s Use of Technology on Their Learning: A Review**  
  *Ching-Ting Hsin, Ming-Chaun Li and Chin-Chung Tsai*  
  85–99

- **Learning of Abstract Concepts through Full-Body Interaction: A Systematic Review**  
  *Laura Malinverni and Narcis Pares*  
  100–116

- **Quality Indicators for Learning Analytics**  
  *Maren Scheffel, Hendrik Drachsler, Slavi Stoyanov and Marcus Specht*  
  117–132

- **Augmented Reality Trends in Education: A Systematic Review of Research and Applications**  
  *Jorge Bacca, Silvia Baldiris, Ramon Fabregat, Sabine Graf and Kinshuk*  
  133–149

- **Closing the Missing Links and Opening the Relationships among the Factors: A Literature Review on the Use of Clicker Technology Using the 3P Model**  
  *Jae Hoon Han*  
  150–168

- **Review of Augmented Paper Systems in Education: An Orchestration Perspective**  
  *Luis P. Prieto, Yun Wen, Daniela Caballero and Pierre Dillenbourg*  
  169–185

- **The Effects of Meta-cognitive Instruction on Students’ Reading Comprehension in Computerized Reading Contexts: A Quantitative Meta-Analysis**  
  *Yi-Chin Lan, Yu-Ling Lo and Ying-Shao Hsu*  
  186–202

### Full Length Articles

- **What Drives Nurses’ Blended e-Learning Continuance Intention?**  
  *Yung-Ming Cheng*  
  203–215

- **Exploring Students’ Knowledge Construction Strategies in Computer-Supported Collaborative Learning Discussions Using Sequential Analysis**  
  *Nurbiha A Shukor, Zaidatun Tasir, Henny Van der Meijden and Jamalludin Harun*  
  216–228

- **The Effect of Using Cooperative and Individual Weblog to Enhance Writing Performance**  
  *H. Gulhan Orhan Karsak, Seval Fer and Feza Orhan*  
  229–241

- **The Role of Group Interaction in Collective Efficacy and CSCL Performance**  
  *Sha-Ling Wang, Hsien-Yuan Hsu, Sunny, S. J. Lin and Gwo-Jen Hwang*  
  242–254

- **Exploring the Moderating Role of Self-Management of Learning in Mobile English Learning**  
  *Rui-Ting Huang*  
  255–267
Development of the Interactive Whiteboard Acceptance Scale (IW BAS): An Initial Study  
*Kung-Teck Wong, Timothy Teo and Pauline Swee Choo Goh*  
268–277

Focused Ubiquity: A Purposeful Approach to Providing Students with Laptops  
*Mike Keppler, Spencer C. Weiler and Dan Maas*  
278–288

Enhancing Students’ NOS Views and Science Knowledge Using Facebook-based Scientific News  
*Hsi-Yu Huan, Huei-Ling Wu, Hsiao-Ching She and Yu-Ren Lin*  
289–301

Development of a Survey to Measure Self-efficacy and Attitudes toward Web-based Professional Development among Elementary School Teachers  
*Chia-Pin Kao, Chin-Chung Tsai and Meilun Shih*  
302–315

A Problem-based Ubiquitous Learning Approach to Improving the Questioning Abilities of Elementary School Students  
*Pi-Hsia Hung, Gwo-Jen Hwang, Yueh-Hsun Lee, Tsung-Hsun Wu, Bahtifar Vogel, Marcelo Milrad and Emil Johansson*  
316–334

Identifying Computer-Supported Collaborative Learning (CSCL) Research in Selected Journals Published from 2003 to 2012: A Content Analysis of Research Topics and Issues  
*Lanqin Zheng, Ronghuai Huang and Junhui Yu*  
335–351

An Augmented Reality-based Mobile Learning System to Improve Students’ Learning Achievements and Motivations in Natural Science Inquiry Activities  
*Tosti H. C. Chiang, Stephen J. H. Yang and Gwo-Jen Hwang*  
352–365

A Multi-modal Digital Game-based Learning Environment for Hospitalized Children with Chronic Illnesses  
*Jui-Chih Chin and Mengping Tsuei*  
366–378

The Nature of Online Charter Schools: Evolution and Emerging Concerns  
*Lisa Haster Waters, Michael K. Barbour and Michael P. Menchaca*  
379–389

Collaborative Inquiry with a Web-Based Science Learning Environment: When Teachers Enact it Differently  
*Daner Sun, Chee-Kit Looi and Wenting Xie*  
390–403

Exploring the Relationship between Self-Regulated Vocabulary Learning and Web-Based Collaboration  
*Sarah Hsueh-Jui Liu, Yu-Ju Lan and Cloudia Ya-Yu Ho*  
404–419

A Study of Learner-Oriented Negative Emotion Compensation in E-learning  
*Jiwei Qin, Qinghua Zheng and Haifei Li*  
420–431

Recognition-Based Physical Response to Facilitate EFL Learning  
*Wu-Yuin Hwang, Timothy K, Shih, Shih-Ching Yeh, Ke-Chien Chou, Zhao-Heng Ma and Worapot Sommool*  
432–445
Studying Research Collaboration Patterns via Co-authorship Analysis in the Field of TeL: The Case of Educational Technology & Society Journal

Panagiotis Zervas\(^1\)\(^2\)*, Asimenia Tzitimdelli\(^1\), Demetrios G. Sampson\(^1\)\(^2\), Nian-Shing Chen\(^3\) and Kinshuk\(^4\)

\(^1\)Department of Digital Systems, University of Piraeus, Greece // \(^2\)Informations Technologies Institute, Centre for Research and Technology Hellas, Greece // \(^3\)National Sun Yat-sen University, Taiwan // \(^4\)Athabasca University, Canada // pzervas@iti.gr // asi77selin@gmail.com // sampson@iti.gr // nianshing@gmail.com // kinshuk@ieee.org

*Corresponding author

**ABSTRACT**

Research collaboration is studied in different research areas, so as to provide useful insights on how researchers combine existing distributed scientific knowledge and transform it into new knowledge. Commonly used metrics for measuring research collaborative activity include, among others, the co-authored publications (concerned with who works with whom) and the citations (concerned with who cites who). Within this context, in this paper, we focus on the co-authorship network of researchers who collaborate in Technology-enhanced Learning (TeL). This is achieved through the example of the Educational Technology & Society (ETS) Journal, where Social Network Analysis (SNA) metrics are applied for analyzing the co-authorship network of the journal. The results of our analysis provided us with evidence that the key authors of ETS Journal co-authorship network have a Taiwanese national background and they have established a strongly connected group that collaborates frequently, diversely and widely. Both authors and collaborations of the ETS co-authorship network have been polynomially increasing during the past 15 years and the co-authorship network has probably arrived at its “phase transition,” which means that it is expected to start becoming more connected in the coming years.

**Keywords**

Research collaboration, Research analytics, Co-authorship network, Social network analysis

**Introduction**

Research collaboration has been explored and discussed in several research areas (Liu et al, 2005; Katz & Martin, 1997; Subramanyam, 1983). Research collaboration can be defined as: “the working together of researchers to achieve the common goal of producing new scientific knowledge” (Katz & Martin, 1997). Commonly used metrics for measuring research collaborative activity include the co-authored publications (concerned with who works with whom) and the citations (concerned with who cites who) (Bukvova, 2010).

On the other hand, research analytics is a research field that utilizes mathematical and algorithmic methods for analysing the way research takes place and it can provide meaningful data for researchers (Harmelen & Workman, 2012). Research analytics are also used for analyzing research collaboration towards identifying meaningful patterns and insights. More specifically, co-authorship has been explored as social networks and it is analyzed by using social network analysis (SNA), whereas citations are analyzed by using bibliometrics (Harmelen & Workman, 2012).

Co-authorship networks fall under an important class of social networks, which are structured through the process of co-authoring a paper between two or more authors (Liu et al., 2005). Co-authorship networks outline the collaboration between researchers and visualize such collaborations in graphs. Analysis of these networks can reveal features of research communities which facilitate in understanding how research collaboration is realized in different research field.

Within this context, co-authorship network analysis has attracted the attention of research community for several research fields. Early research works have focused on defining elements (e.g., network structure and typology, as well as positioning of certain authors) that are meaningful to be analyzed within these networks (Moody, 2004; Newman, 2001; Barabási & Albert, 1999). Recent research works have mainly focused on analyzing these elements within specific co-authorship networks that are developed by papers co-authored for proceedings of scientific conferences or for scientific journals addressing various research fields (Erfanmanesh et al., 2012; Erman & Todorovski, 2010; Velden et al., 2010; Yan & Ding, 2010). In the field of Technology enhanced Learning (TeL), there is some studies that aim to analyze co-authorship networks (Pham et al., 2012; Ochoa et al., 2009; Kienle, & Wessner, 2005). However, we can notice that these studies are limited and they have been mainly based on studying...
conferences related to the TeL research field. Thus, one issue to investigate is to study the co-authorship networks developed by prominent scientific journals, which include more mature research results and they could reflect the current state of research collaboration realized in TeL in a more robust way.

To this end, in this paper we focus on the co-authorship network of researchers who have published their joint work in the Educational Technology & Society (ETS) Journal from 1999 to 2012. The purpose of our study is to utilize SNA metrics to reveal useful patterns and features within this co-authorship network, as well as to identify the expected evolution of this network. The results of our study could provide a first step towards gaining a thorough insight on how research collaboration is realized in TeL.

The paper is structured as follows. Following this introduction, the next section presents background information about how co-authorship network analysis can be achieved by using SNA metrics, namely graph metrics and vertex specific metrics and we present those metrics exploited in our study. Afterwards, we discuss related works from the literature that deal with co-authorship networks analysis in the field of TeL. Next, we present the method of analysis of the co-authorship network developed by the ETS Journal and we present the results of this analysis. Finally, we present our conclusions and ideas for future work.

**Background: Co-authorship network analysis**

Co-authorship network analysis can be achieved by using SNA metrics (Otte & Rousseau, 2002). These metrics are used to identify features and patterns of: (a) the typology of the co-authorship network; and, (b) the centralities of the authors of the network. The former can be addressed by graph metrics (as presented in Table 1), whereas the latter can be addressed by vertex specific metrics (as presented in Table 2).

**Table 1. Graph metrics in co-authorship networks**

<table>
<thead>
<tr>
<th>SNA Metric</th>
<th>Definition</th>
<th>Application in co-authorship networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Edges</td>
<td>It is used to measure the number of edges, which occur exactly once between two vertices</td>
<td>If combined with the number of vertices (namely, authors), it can denote the average number of co-authors of the publications modelled by the network</td>
</tr>
<tr>
<td>Edges with Duplicates</td>
<td>It is used to measure edges that occur more than once between two vertices</td>
<td>It counts the number of repeating collaborations between authors</td>
</tr>
<tr>
<td>Total Edges</td>
<td>It is the sum of unique and duplicate edges in the graph</td>
<td>It counts all collaborations that exist in the network</td>
</tr>
<tr>
<td>Connected Components</td>
<td>It identifies and measures closed subgroups of vertices that are interconnected, but they do not connect with other vertices or groups outside of the specific subgroups</td>
<td>It denotes the number of closed subgroups of authors, which are not collaborating with other specific subgroups</td>
</tr>
<tr>
<td>Maximum Vertices in a Connected Component</td>
<td>It measures the maximum number of vertices that are connected to each other in a component. This component is called “Giant component”</td>
<td>It denotes the number of authors that can be found in the biggest subgroup</td>
</tr>
<tr>
<td>Maximum Edges in a Connected Component</td>
<td>It measures the number of edges that exist inside the “Giant component”</td>
<td>It counts the number of collaborations that can be found in the biggest subgroup</td>
</tr>
<tr>
<td>Maximum Geodesic Distance (Diameter)</td>
<td>It is used to measure the shortest possible distance between the two farthest vertices in the network.</td>
<td>It denotes the maximum distance that a research idea will need, in order to be spread between the two farthest authors in the network</td>
</tr>
<tr>
<td>Average Geodesic Distance</td>
<td>It is used to measure how close in average two vertices are placed, inside the graph of the social network</td>
<td>It denotes the average distance that a research idea will need, in order to be spread between any two authors in the network</td>
</tr>
</tbody>
</table>
Graph Density  
It represents the completeness of vertices’ interconnections. The maximal density is 1 (for complete graphs) and the minimal density is 0  
It denotes the degree of collaboration that takes place within the network

Modularity  
It is used to measure the degree of clustering that appears in the graph. This means that if a graph consists of distinct closed subgroups, then it has bigger modularity. Modularity is measured by comparing the number of edges inside a subgroup with the number of edges in the whole network  
It denotes the degree of fragmentation of the network to closed subgroups of authors

<table>
<thead>
<tr>
<th>SNA Metric</th>
<th>Definition</th>
<th>Application in co-authorship networks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Degree Centrality</strong></td>
<td>It is used to measure the “popularity” of a vertex. This means that it measures the number of unique edges that create a connection between this specific vertex and others. Each connection is only counted once, even if it occurs multiple times.</td>
<td>It denotes the existence of authors that collaborate very often with many other authors</td>
</tr>
<tr>
<td><strong>Betweenness Centrality</strong></td>
<td>It is based on the number of shortest paths passing through a vertex. Vertices with high betweenness centrality play the role of connecting different subgroups of the social network.</td>
<td>It denotes the existence of authors that connect and strengthen the collaboration between different subgroups of authors</td>
</tr>
<tr>
<td><strong>Closeness Centrality</strong></td>
<td>It represents the distance of a vertex to all other in the network by focusing on the geodesic distance from each vertex to all others. Closeness centrality can be interpreted as a metric of how long it will take information to spread from a given vertex to others in the network.</td>
<td>It denotes the existence of authors that have collaborated with researchers from a wide variety of research areas and they can quickly spread research ideas across the network.</td>
</tr>
<tr>
<td><strong>Eigenvector Centrality</strong></td>
<td>It is used to measure the degree of a vertex and the degree of the vertices with which this vertex is connected. Eigenvector centrality can identify vertices that are connected with other “popular” vertices.</td>
<td>It denotes the existence of authors that are more likely to receive first new research ideas that are spread across the network.</td>
</tr>
</tbody>
</table>

**Table 2. Vertex specific metrics in co-authorship networks**

Related studies: **Co-authorship network analysis in TeL**

In this section, we provide an overview of existing studies that focus on co-authorship network analysis in TeL. Kienle & Wessner (2005) have studied the co-authorship network of the researchers that publish their papers in the proceedings of “Computer Supported Collaborative Learning (CSCL)” conference. Within their study, the authors have focused on the evolution of the co-authorship network and the identification of key authors. Their findings revealed a relatively small number of recurring participants and a growing tendency for international participation.

Ochoa et al. (2009) have analyzed the co-authorship network of the researchers that publish their papers to the proceedings of the “World Conference on Educational Media and Technology (ED-MEDIA).” The study focused on the identification of influential authors by applying vertex specific centrality metrics for authors’ ranking and the identification of important groups of the network. The analysis reveals that ED-MEDIA is a vibrant and collaborative community. The results are also used to propose a personalized recommender system for future participants.
Reinhardt et al. (2011) have analyzed the co-authorship network of the researchers that publish their papers to the proceedings of the European conference on technology enhanced learning (EC-TEL). Within their study, the authors have identified important groups of the network and they have indicated that EC-TEL is a highly fragmented conference which consists of a large number of weakly connected subgroups. Moreover, their findings showed a very fragmented co-authorship network, with a small number of researchers with high centralities.

Pham et al. (2012) studied evolution patterns of co-authorship networks for five TeL conferences, namely ICALT, ECTEL, ICWL, ITS, and AIED. Thus, they proposed a development model with different stages, which can be applied in order to determine in which stage a conference is. The authors have also applied SNA metrics to the co-authorship networks of these five TeL conferences and they have compared the results with four conferences in the Database research field. According to their analysis, all five conferences follow the same development stages, and it appears that more interdisciplinary conferences have a slower development rate.

Finally, Cho et al. (2013) have conducted a citation network analysis of Educational Technology Research and Development (ETR&D) journal to examine the trends and issues of the educational technology field’s scholarly community that have evolved in the past two decades. The authors have used SNA metrics to the citation network of the ETR&D journal to identify influential papers and researchers in the field. Their findings showed that frequently co-cited papers were recognized as having a strong relationship by a few researchers and five cohesive subgroups of authors generated key research areas in the field.

As we can notice from the aforementioned studies, analysis of co-authorship networks can be useful for studying research collaboration in the field of TeL. This means that influential authors can be identified, as well as important groups of authors that work on specific research areas in the field of TeL. Nevertheless, most existing studies have been focused on co-authorship networks developed by researchers who publish their papers to the proceedings of international conferences, whereas the study conducted by Cho et al. (2013) mainly focused on analyzing the citation network of a journal and not its co-authorship network. Thus, one issue to investigate is the co-authorship networks developed by highly ranked scientific journals. The research presented in this paper addresses this issue through the example of the Educational Technology & Society (ETS) Journal where SNA metrics are applied for analyzing the co-authorship network of the journal.

**Method of analysis**

**Sample**

The ETS Journal was selected because of the following three reasons: (a) it is open access and consequently we were able to have access to all published papers; (b) it is an accredited journal because it has an impact factor of 1.171 according to Thomson Scientific 2012 Journal Citations Report and it is currently ranked 4th in the top-20 publications for Educational Technology in Google Scholar (http://goo.gl/NVyDpb); and, (c) it is published for more than 15 consecutive years, since 1998.

The papers were collected from ETS Journal website (http://www.ifets.info/). More specifically, papers from 15 volumes were collected that correspond to 57 issues. These issues span a time period of 14 years, from January 1999 to December 2012, which is translated into 1041 papers, namely 589 regular papers and 452 special issues papers. More information about the theme of special issue papers included in our sample can be found at the ETS Journal website (http://www.ifets.info/issues.php?show=sp_issues). All single author papers (249 papers) were excluded from the sample, since the scope of our study is to study the co-authorship network of the ETS Journal. Consequently, we considered in our sample 455 regular papers and 337 special issues papers, which were co-authored by two or more authors.

**Process**

From the types of papers previously mentioned, a list of authors who have jointly authored papers was created with the use of NodeXL (http://nodexl.codeplex.com/). In this list, the year of publication, as well as the keywords for
every paper, was also included. By following the above process, we created the ETS Journal co-authorship network, where authors are represented as vertices and co-authorship connections are represented as edges. Figure 1 provides an overview of the ETS Journal co-authorship graph of the network.

In Figure 1, the visual representation of the network graph has been optimized by placing small groups at the bottom of the graph, so as to focus on bigger groups which formulate the core of the co-authorship network.

Educational Technology & Society (ETS) Journal Co-authorship network analysis

In this section, we demonstrate the application of SNA metrics, which have been previously presented to the ETS Journal Co-Authorship Graph (as presented in Figure 1). Our analysis consists of two stages, namely (a) Network Identity: at this stage, we analyze the identity of the current state of the network by applying graph metrics and we perform a ranking and a clustering of authors by applying vertex specific centrality metrics and (b) Network Evolution: at this stage, we analyze the annual evolution of the network by monitoring the changes of specific graph metrics. The type of analysis that is conducted in our network is exploratory. This means that there is not a specific hypothesis to test but the conclusions emerge from the analysis itself (de Nooy et al., 2005).

Co-authorship network identity

Graph metrics

Table 3 presents the values of the graph metrics that have been applied to the ETS Journal co-authorship network. These values have been calculated by importing our data to NodeXL.
As we can notice from Table 3, there are 1944 collaborating authors in this network and 2776 unique collaborations. This means that on average an author collaborates with 1.43 other authors. Thus, having excluded the single author papers, we can mention that, on average, the authoring team of an ETS Journal paper consists of 2 or 3 authors. These are slightly lower values comparing to co-authorship networks from other research fields such as biology and physics constructed by Newman (2001), who has found that authors per paper for biology co-authorship network are 3.75 and for physics co-authorship network, authors per paper are 2.53. This provides us with indications that TeL researchers/authors are collaborating in smaller groups compared to other research fields.

Additionally, we can notice that there are 576 duplicated collaborations that correspond to 17.18% of the total collaborations. This means that only 1 out of 6 collaborations have occurred twice or more times. An important issue is the high number of connected components (468) in a network of 1944 vertices. This means that this network includes many subgroups that are collaborating independently from the main core of collaborations (namely, the giant component). This claim is also supported by the modularity metric, which is 0.868. An additional metric that supports this claim is the graph density, which is very low (0.002). This also means that the authors included in the various subgroups of the network are rarely collaborating with authors outside from these subgroups.

Another interesting element is the giant component of the network, which includes 226 authors and 770 collaborations. These values correspond to 11.63% of the total authors and to 22.97% of the total collaborations. This means that authors included in the giant component collaborate more often than the authors who are not part of the giant component.

Other important metrics are the diameter of the network and the average distance of the network. More precisely, the diameter of the network is 9. This value appears to be the same with the diameter of the mathematics co-authorship networks studied by Barabási et al. (2002), as well as with the library and information science co-authorship networks studied by Yan & Ding (2010). This means that the co-authorship network’s typology in TeL follows similar patterns with other research fields. Finally, the average distance of the network highlights that 3,766 steps are needed to reach an author within the network. The values of these metrics reveal that research ideas within the network are not spreading quickly.

**Vertex specific centrality metrics**

As it is clearly identified from Figure 1, as well as from the graph metrics presented in the previous paragraph, the ETS journal co-authorship network includes many subgroups that are not connected to each other. As a result, centrality metrics (such as betweenness centrality, closeness centrality and eigenvector centrality) cannot be applied to the whole network because these metrics are calculated by taking into account the distance between vertices, which is infinite for disconnected subgroups (Opsahl et al., 2010). In order to overcome this problem, we applied the centrality metrics only to the giant component of the network. Figure 2 presents the distribution of the calculated centrality metrics for the authors of the giant component.
The frequency of degree centrality (Figure 2a), betweeness centrality (Figure 2b) and eigenvector centrality (Figure 2d) follows power-law distribution where most authors have low centrality values while a few authors have high degree centrality values. On the other hand, the distribution of closeness centrality (Figure 2c) follows the normal curve distribution. These distributions appear to be similar with frequency distributions of centralities in other research fields such as mathematics and neuro-science co-authorship networks studied by Barabási et al. (2002), library and information science co-authorship networks studied by Yan & Ding (2010) and in e-Government co-authorship networks studied by Ernan & Todorovski (2010). This provides us with indications that TeL researchers/authors in ETS Journal collaborate by following similar patterns with other research fields.

In order to analyze insights of the central authors of the ETS co-authorship, we have calculated the top 30 authors based on degree centrality, betweenness centrality, closeness centrality and eigenvector centrality (Table 4). Authors appear in the top 30 of all centrality metrics are marked in bold and italic fonts.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Degree Centrality</th>
<th>Value</th>
<th>Betweenness Centrality</th>
<th>Value</th>
<th>Closeness Centrality</th>
<th>Value</th>
<th>Eigenvector Centrality</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gwo-Jen Hwang</td>
<td>36</td>
<td>Kinshuk</td>
<td>8629,879</td>
<td>Nian-Shing Chen</td>
<td>0,00165</td>
<td>Kuo-En Chang</td>
<td>0,071</td>
</tr>
<tr>
<td>2</td>
<td>Yueh-Min Huang</td>
<td>31</td>
<td>Tak-Wai Chan</td>
<td>7230,966</td>
<td>Stephen Yang</td>
<td>0,00157</td>
<td>Yao-Ting Sung</td>
<td>0,070</td>
</tr>
<tr>
<td>3</td>
<td>Chin-Chung Tsai</td>
<td>26</td>
<td>Nian-Shing Chen</td>
<td>6965,082</td>
<td>Kinshuk</td>
<td>0,00155</td>
<td>Chun-Hua Chen</td>
<td>0,062</td>
</tr>
<tr>
<td>4</td>
<td>Nian-Shing Chen</td>
<td>22</td>
<td>Stephen Yang</td>
<td>5293,760</td>
<td>Gwo-Jen Hwang</td>
<td>0,0015</td>
<td>Yun-Cheng Luo</td>
<td>0,062</td>
</tr>
<tr>
<td>5</td>
<td>Kuo-En Chang</td>
<td>22</td>
<td>Gwo-Jen Hwang</td>
<td>5145,543</td>
<td>Yueh-Min Huang</td>
<td>0,00148</td>
<td>Chen-Jung Huang</td>
<td>0,062</td>
</tr>
<tr>
<td>6</td>
<td>Yao-Ting Sung</td>
<td>21</td>
<td>Wu-Yuin Hwang</td>
<td>5115,723</td>
<td>Chin-Chung Tsai</td>
<td>0,00146</td>
<td>Ming-Chou Liu</td>
<td>0,060</td>
</tr>
<tr>
<td>7</td>
<td>Tak-Wai Chan</td>
<td>20</td>
<td>Kuo-En Chang</td>
<td>4838,570</td>
<td>Wu-Yuin Hwang</td>
<td>0,00145</td>
<td>Tz-Hau Huang</td>
<td>0,059</td>
</tr>
<tr>
<td>8</td>
<td>Wu-Yuin Hwang</td>
<td>20</td>
<td>Yueh-Min</td>
<td>4678,167</td>
<td>Tak-Wai</td>
<td>0,00139</td>
<td>Hung-Yen</td>
<td>0,059</td>
</tr>
<tr>
<td>Rank</td>
<td>Author Name</td>
<td>Degree Score</td>
<td>Pageviews</td>
<td>Citations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Stephen Yang</td>
<td>3354,892</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Shian-Shyong Tseng</td>
<td>3223,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Kinshuk</td>
<td>2349,809</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Chun-Hua Chen</td>
<td>2131,545</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Yun-Cheng Luo</td>
<td>1760,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Chen-Jung Huang</td>
<td>1736,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Ming-Chou Liu</td>
<td>1637,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Chin-Yeh Wang</td>
<td>1579,189</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Hui-Chun Chu</td>
<td>1538,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Ting-Ting Wu</td>
<td>1532,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Tz-Hau Huang</td>
<td>1526,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Hung-Yen Shen</td>
<td>1401,532</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Kuo-Liang Huang</td>
<td>1100,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Jia-Jian Liao</td>
<td>887,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Kai-Wen Hu</td>
<td>871,267</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Tun-Yu Chang</td>
<td>775,803</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Ben Chang</td>
<td>710,495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Gwo-Dong Chen</td>
<td>668,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Tzu-Chien Liu</td>
<td>666,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Chee-Kit Looi</td>
<td>666,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Demetrios Sampson</td>
<td>536,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Jun-Ming Su</td>
<td>447,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first column of Table 4 shows the ranking of the top 30 authors based on the degree score. The top 3 authors in terms of degree centrality are: Gwo-Jen Hwang (15 papers, National Taiwan University of Science and Technology, Taiwan), Yueh-Min Huang (15 papers, National Cheng Kung University, Taiwan) and Chin-Chung Tsai (13 papers, National Taiwan University of Science and Technology, Taiwan). These authors are collaborating frequently with many other authors and they can be viewed as popular authors.

With regard to betweeness centrality (second column of Table 4), the top 3 authors are: Kinshuk (8 papers, Athabasca University, Canada) followed by Tak-Wai Chan (5 papers, National Central University, Taiwan) and Nian-Shing Chen (11 papers, National Sun Yat-sen University, Taiwan). Thus, these authors connect and strengthen...
collaboration between different groups of authors and with the giant component and they can be viewed as central authors in the ETS Journal co-authorship network.

In terms of closeness centrality (third column of Table 4), the scores of the top 30 authors were very close with the leaders being Nian-Shing Chen (11 papers, National Sun Yat-sen University, Taiwan) followed by Stephen Yang (7 papers, National Central University, Taiwan) and Kinshuk (8 papers, Athabasca University, Canada). These authors have collaborated with researchers from a wide range of research areas in the field of TeL and this means that they can be viewed as influential authors, who can quickly spread research ideas across the network.

Finally, the last column of Table 4 shows the ranking of the top 30 authors based on the eigenvector score. The top 3 authors in terms of eigenvector centrality are: Kuo-En Chang (7 papers, National Taiwan Normal University, Taiwan), Yao-Ting Sung (7 papers, National Taiwan Normal University, Taiwan) and Chun-Hua Chen (2 papers, National Dong Hwa University, Taiwan). These authors are connected to many other authors who are well connected and popular and thus they are more likely to co-author new papers with those authors and are directly introduced to new concepts or research ideas.

It is also important to mention that two of the top three positions regarding betweenness centrality and closeness centrality have been taken by the two out of three editors of the ETS Journal, namely Kinshuk and Nian-Shing Chen. This means that ETS editors play an important role to the collaborations that are developed within the ETS Journal co-authorship network, as they connect various sub-groups (as identified by the betweenness centrality) and they can receive research ideas spread within the network very quickly (as identified by the closeness centrality). Moreover, another interesting fact of the top 3 authors according to the different centrality metrics is that all of them are from Taiwan except from Kinshuk, who is from Canada. This means that the ETS co-authorship network includes a strongly connected subgroup of authors, who are from Taiwan and they collaborate frequently (as indicated by degree centrality), diversely (as indicated by betweenness centrality) and widely (as indicated by closeness and eigenvector centrality). This claim is also supported by the country of origin of the authors that appear in the top 30 of all centrality metrics and they are marked in bold and italic fonts. These authors are: Yueh-Min Huang (National Cheng Kung University, Taiwan), Kuo-En Chang (National Taiwan Normal University, Taiwan), Tak-Wai Chan (National Central University, Taiwan) and Wu-Yuin Hwang (National Central University, Taiwan). All these authors are from Taiwan but come from various universities. This gives an indication that research collaboration between Taiwanese researchers from different universities and/or research centres is probably promoted and facilitated by national research policy in Taiwan.

In order to identify whether authors’ ranking based on the centrality metrics can play an important role to the recognition of their research work, we studied the correlation of the centrality metrics of the top 30 authors with the number of citations to their publications. Table 5 presents the top 30 author per centrality metric, as well as their citations. It is worth mentioning that the citations of top 30 authors were calculated form Google Scholar service (http://scholar.google.gr/) and these are the numbers that were available on December 23, 2013.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Degree Centrality</th>
<th>Value</th>
<th>Citations Number</th>
<th>Betweenness Centrality</th>
<th>Value</th>
<th>Citations Number</th>
<th>Closeness Centrality</th>
<th>Value</th>
<th>Citations Number</th>
<th>Eigenvector Centrality</th>
<th>Value</th>
<th>Citations Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gwo-Jen Hwang</td>
<td>36</td>
<td>3393</td>
<td>Kinshuk</td>
<td>8629,879</td>
<td>3676</td>
<td>Nian-Shing Chen</td>
<td>Stephen Yang</td>
<td>0.00165</td>
<td>1761</td>
<td>Kuo-En Chang</td>
<td>0.071</td>
</tr>
<tr>
<td>2</td>
<td>Yueh-Min Huang</td>
<td>31</td>
<td>4999</td>
<td>Tak-Wai Chan</td>
<td>7230,966</td>
<td>2848</td>
<td>Nian-Shing Chen</td>
<td>Stephen Yang</td>
<td>0.00157</td>
<td>3226</td>
<td>Yao-Ting Sung</td>
<td>0.070</td>
</tr>
<tr>
<td>3</td>
<td>Chin-Chung Tsai</td>
<td>26</td>
<td>7321</td>
<td>Nian-Shing Chen</td>
<td>6965,082</td>
<td>1761</td>
<td>Kinshuk</td>
<td></td>
<td>0.00155</td>
<td>3676</td>
<td>Chun-Hua Chen</td>
<td>0.062</td>
</tr>
<tr>
<td>4</td>
<td>Nian-Shing Chen</td>
<td>22</td>
<td>1761</td>
<td>Stephen Yang</td>
<td>5293,760</td>
<td>3393</td>
<td>Gwo-Jen Hwang</td>
<td></td>
<td>0.0015</td>
<td>3393</td>
<td>Yun-Cheng Luo</td>
<td>0.062</td>
</tr>
<tr>
<td>No.</td>
<td>Name</td>
<td>Age</td>
<td>ID</td>
<td>Gwo-Jen Hwang Wu-Yuin Hwang Kuo-En Chang</td>
<td>Yueh-Min Huang Chin-Chung Tsai Wu-Yuin Hwang Tak-Wai Chan Tsu-Chien Liu Irene Chen</td>
<td>0.00148</td>
<td>4999</td>
<td>Chen-Shing Jung Huang Ming-Chou Liu Tz-Hau Huang</td>
<td>0.062</td>
<td>614</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------</td>
<td>-----</td>
<td>-------</td>
<td>-----------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>--------</td>
<td>-----</td>
<td>------------------------------------------------</td>
<td>--------</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Kuo-En Chang</td>
<td>22</td>
<td>1620</td>
<td>5145,543 510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Yao-Ting Sung</td>
<td>21</td>
<td>1375</td>
<td>5115,723 3226</td>
<td></td>
<td>0.00146</td>
<td>7321</td>
<td></td>
<td>0.060</td>
<td>225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Tai-Wai-Chen-Wu-Yuin</td>
<td>20</td>
<td>2848</td>
<td>4838,570 1620</td>
<td></td>
<td>0.00145</td>
<td>510</td>
<td></td>
<td>0.059</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Huang-Stephen-Yang</td>
<td>16</td>
<td>3226</td>
<td>4678,167 4999</td>
<td></td>
<td>0.00139</td>
<td>2848</td>
<td></td>
<td>0.059</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Shian-Shyong-Tseng</td>
<td>15</td>
<td>2185</td>
<td>3354,892 7321</td>
<td></td>
<td>0.00139</td>
<td>808</td>
<td></td>
<td>0.059</td>
<td>393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Kinshuk</td>
<td>14</td>
<td>3676</td>
<td>3223,000 1829</td>
<td></td>
<td>0.00138</td>
<td>507</td>
<td></td>
<td>0.059</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Chun-Hua-Chen-Yun-Cheng-Luo</td>
<td>13</td>
<td>614</td>
<td>2349,809 496</td>
<td></td>
<td>0.00138</td>
<td>217</td>
<td></td>
<td>0.059</td>
<td>107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Shyan-Shyong-Tseng</td>
<td>13</td>
<td>225</td>
<td>1736,000 4353</td>
<td></td>
<td>0.00136</td>
<td>225</td>
<td></td>
<td>0.018</td>
<td>162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Chin-Yeh-Hsi-Chen-Chu</td>
<td>12</td>
<td>774</td>
<td>1579,189 225</td>
<td></td>
<td>0.00134</td>
<td>357</td>
<td></td>
<td>0.015</td>
<td>156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Hui-Chung-Chu-Ting-Ting-Wu</td>
<td>12</td>
<td>2776</td>
<td>1559,000 709</td>
<td></td>
<td>0.00133</td>
<td>406</td>
<td></td>
<td>0.015</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Tz-Hau-Huang</td>
<td>11</td>
<td>23</td>
<td>1526,000 2006</td>
<td></td>
<td>0.00133</td>
<td>32</td>
<td></td>
<td>0.014</td>
<td>1207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Hung-Yen-Shen-Liang-Huang-Jia-Jian-Liao</td>
<td>11</td>
<td>31</td>
<td>1401,532 406</td>
<td></td>
<td>0.00132</td>
<td>39</td>
<td></td>
<td>0.013</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Kuo-Liang-Huang</td>
<td>11</td>
<td>393</td>
<td>1100,000 132</td>
<td></td>
<td>0.00132</td>
<td>45</td>
<td></td>
<td>0.013</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Huang-Jia-Jian-Liao</td>
<td>11</td>
<td>37</td>
<td>887,000 2355</td>
<td></td>
<td>0.00132</td>
<td>1620</td>
<td></td>
<td>0.013</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Kai-Wen-Hu</td>
<td>11</td>
<td>107</td>
<td>871,267 1375</td>
<td></td>
<td>0.00128</td>
<td>370</td>
<td></td>
<td>0.013</td>
<td>201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Tun-Yu-Chang-Ben</td>
<td>11</td>
<td>29</td>
<td>775,803 446</td>
<td></td>
<td>0.00126</td>
<td>737</td>
<td></td>
<td>0.012</td>
<td>493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Shui-Cheng-Lin-Yu-Ru-Hong</td>
<td>11</td>
<td>2171</td>
<td>710,495 2776</td>
<td></td>
<td>0.00126</td>
<td>200</td>
<td></td>
<td>0.008</td>
<td>510</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
According to the data presented in Table 5, we calculated the Pearson’s Correlation Coefficient between number of citations and centrality metrics. The results are presented in Table 6.

Table 6. Correlation between number of citation and centrality metrics

<table>
<thead>
<tr>
<th>Centrality Metrics</th>
<th>Degree Centrality</th>
<th>Betweenness Centrality</th>
<th>Closeness Centrality</th>
<th>Eigenvector Centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.60</td>
<td>0.44</td>
<td>0.62</td>
<td>-0.20</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level.

As we can notice from Table 6, three centrality metrics (namely, degree centrality, betweenness centrality and closeness centrality) have significant strong correlation with number of citations at the 0.01 level, with closeness centrality as the highest. On the other hand, there is no correlation between number of citations and eigenvector centrality. As a result, we can claim that degree centrality, betweenness centrality and closeness centrality can possibly be used as supplementary indicators for assessing author’s scientific recognition, providing alternative perspectives to the current methods, such as h-index and i10-index.

Authors’ clustering and key research areas of collaboration

In this section and by using NodeXL’s clustering algorithm, which clusters the vertices of a social network based on how they are connected, we perform a clustering of the authors included in the ETS journal co-authorship network. The scope of this process is to identify major clusters of authors, as well as the papers that they have co-authored towards identifying key research areas of collaboration, of the papers published to ETS Journal. In order to achieve this, we focused on the papers’ keywords of each cluster and we have used a word count tool (http://www.writewords.org.uk/word_count.asp) for measuring the frequency of keywords in the papers of each cluster. This will facilitate us to identify relevant research areas for each cluster. Table 7 presents the four (4) major authors’ clusters that have been identified. It is worth mentioning that we have focused on those clusters that include at least 20 papers (approximately 2.50 % of our sample), in order to have a substantial number of keywords to analyze. Clusters after cluster 4 include less than 6 papers and for this reason, they were not analyzed.

Table 7. ETS Journal major authors’ clusters

<table>
<thead>
<tr>
<th>No</th>
<th>Authors (total authors: 1944)</th>
<th>Papers (total papers: 792)</th>
<th>Top 3 Frequent Papers’ Keywords</th>
<th>Total Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>162 (8%)</td>
<td>85 (11%)</td>
<td>mobile learning ubiqutious learning collaborative learning</td>
<td>24</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>83 (4%)</td>
<td>30 (4%)</td>
<td>learning design / educational modelling / IMS learning design learning object metadata</td>
<td>9</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>46</td>
<td>22</td>
<td>computer supported collaborative learning</td>
<td>4</td>
</tr>
</tbody>
</table>
As we can notice from Table 5, cluster 1 appears to focus on “Wireless, Mobile and Ubiquitous Technologies for Learning” according to the top 3 frequent keywords identified in its papers with 48 out of 85 papers related to this research area. Cluster 2 appears to focus on “International Standards and Specifications for Learning Technologies” with 21 out of 30 papers related to this research area. Cluster 3 appears to focus on “Computer Supported Collaborative Learning” and Cluster 4 on “Adaptive and Personalized Technology-Enhanced Learning,” with 10 out of 22 and 15 out of 21 papers respectively. These research areas have also been identified by Kinshuk et al. (2013) as main research areas of the ETS Journal, when analyzing of the topics of highly cited papers published in the ETS Journal for the period from 2003 to 2010. This means that established clusters within the ETS Journal co-authorship network with an adequate number of published papers can attract high numbers of citations and thus become influential in TeL research community.

Co-authorship network evolution

In this section, we study the evolution of the ETS Journal co-authorship network. With this process, we can understand the current status of the network and we can also make predictions about the future status of the network. Figure 3 and Figure 4 present the annual accumulative evolution of ETS authors and collaborations respectively.
As we can notice from Figure 6 and 7 the number of authors and collaborations increases gradually. The two curves fit $y = 4.0055x^2 + 87.287x - 50.121$ and $y = 14.189x^2 + 42.417x + 23.104$ respectively, with coefficients of determination $R^2 = 0.9973$ and $R^2 = 0.9971$. This result indicates that there is a growing tendency for authors and collaborations, which is expected to increase approximately with these curves in the coming years.

Table 8 presents the properties of the evolving ETS co-authorship network from 1999 to 2012. It includes information about the annual accumulative distribution of authors and collaborations for the ETS co-authorship network, as well as for its giant component. Moreover, for the giant component, we have calculated the annual ratio of its authors per annual total number of authors, as well as the annual ratio of its collaborations per annual total number of collaborations. Finally, we have calculated the annual ratio of collaborations for the entire network and the annual average distance for the ETS co-authorship network and its giant component.

<table>
<thead>
<tr>
<th>Year</th>
<th>Accumulative Number of Authors</th>
<th>Accumulative Number of Collab.</th>
<th>Ratio of Collab.</th>
<th>Avg. Distance</th>
<th>Giant component</th>
<th>Total</th>
<th>Giant component</th>
<th>Total</th>
<th>Giant component</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>25</td>
<td>24</td>
<td>0.96</td>
<td>0.76</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>2000</td>
<td>188</td>
<td>230</td>
<td>1.22</td>
<td>0.76</td>
<td>6</td>
<td>3.19%</td>
<td>9</td>
<td>3.91%</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>243</td>
<td>304</td>
<td>1.25</td>
<td>0.76</td>
<td>12</td>
<td>4.94%</td>
<td>14</td>
<td>4.61%</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>374</td>
<td>459</td>
<td>1.23</td>
<td>0.79</td>
<td>16</td>
<td>4.28%</td>
<td>31</td>
<td>6.75%</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>471</td>
<td>592</td>
<td>1.26</td>
<td>0.84</td>
<td>21</td>
<td>4.46%</td>
<td>37</td>
<td>6.25%</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>575</td>
<td>734</td>
<td>1.28</td>
<td>0.93</td>
<td>33</td>
<td>5.74%</td>
<td>61</td>
<td>8.31%</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>709</td>
<td>937</td>
<td>1.32</td>
<td>1.05</td>
<td>41</td>
<td>5.78%</td>
<td>74</td>
<td>7.90%</td>
<td>1.61</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>917</td>
<td>1267</td>
<td>1.38</td>
<td>1.26</td>
<td>56</td>
<td>6.11%</td>
<td>108</td>
<td>8.52%</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>1067</td>
<td>1502</td>
<td>1.41</td>
<td>1.46</td>
<td>71</td>
<td>6.65%</td>
<td>131</td>
<td>8.72%</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>1262</td>
<td>1882</td>
<td>1.49</td>
<td>2.69</td>
<td>125</td>
<td>9.90%</td>
<td>328</td>
<td>17.43%</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>1455</td>
<td>2318</td>
<td>1.59</td>
<td>2.97</td>
<td>153</td>
<td>10.52%</td>
<td>438</td>
<td>18.90%</td>
<td>3.74</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>1567</td>
<td>2657</td>
<td>1.70</td>
<td>3.63</td>
<td>170</td>
<td>10.85%</td>
<td>595</td>
<td>22.39%</td>
<td>4.37</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>1732</td>
<td>2921</td>
<td>1.69</td>
<td>3.72</td>
<td>194</td>
<td>11.20%</td>
<td>653</td>
<td>22.36%</td>
<td>4.41</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>1944</td>
<td>3352</td>
<td>1.72</td>
<td>3.76</td>
<td>226</td>
<td>11.63%</td>
<td>769</td>
<td>22.94%</td>
<td>4.27</td>
<td></td>
</tr>
</tbody>
</table>

As we can notice from Table 8, there is an increase of the ratio of collaborations per year. This means that as the journal and the field, in general, matures, authors collaborate more widely. Moreover, ratio of giant component’s authors per total authors, as well as the ratio of giant component’s collaborations per total collaborations increases per year. This provides us with indications that small groups of authors start collaborating more widely in recent years and they enter the giant component. As a result, we could expect in the coming years that the co-authorship network will become from highly fragmented to adequate connected. Finally, we can notice that the average distance of both the network and the giant component increases per year. This is due to the fact that more new authors are publishing papers to the ETS Journal each year. However, if we focus on the time period from 2010 to 2012, we can notice that there is a slight increase of network’s average distance. On the contrary, the giant component’s average distance was slightly increased from 2010 to 2011 and slightly decreased from 2011 to 2012. This provides us with indications that the network might have arrived at its “phase transition” (Lee et al., 2010) in 2010, which is the phase where authors are starting collaborating with each other much more frequently and more widely. Based on that, we can also assume that the average distance of both the network and the giant component will start decreasing in the coming years.

Another issue that we investigated was the development model of the ETS Journal co-authorship network. As it has been identified by the Pharm et al. (2012) co-authorship networks follow a specific development pattern that consists of four stages, as follows:

- **Born Phase**: this is the initial phase of the co-authorship network, where there are only a few connections between authors.
- **Bonding Phase**: this is the phase after some years, when author groups start becoming visible in the network.
- **Emergence Phase**: this is the phase, when authors groups created during the bonding phase are gradually integrated through publications that involve authors from more than one group.

- **Stable Phase**: this is the phase, when the co-authorships network has formed a specific type of topology. There are different types of topologies that include: (a) focused topology, that is for networks featuring a strongly connected core group of authors that is connected to other smaller groups, (b) interdisciplinary topology, that is for networks featuring several groups connected via some gatekeepers, but where there is no core group and (c) hierarchical topology, that is for networks featuring some “super gatekeepers” who connect a hierarchy of groups.

Figure 5 presents the development of the ETS Journal co-authorship network in four time periods, namely 1999 to 2003 (5 years), 1999 to 2007 (9 years), 1999 to 2010 (12 years) and 1999 to 2012 (14 years). As we can notice from Figure 5, in 2003 there were many small groups formed by authors who have co-authored one paper with at least one co-author. Four years later, in 2007, we can notice the existence of some larger co-authorship groups, and also some bonding among author groups. In 2010, some larger co-authorship groups are already clearly discrete, and the network also starts to form a clearly visible large component of core authors in the emergence phase. By 2012, the largest component is beginning to actually deserve the label “giant component” and we can notice that many members of the giant component have co-authorship ties to other authors but author groups at the periphery remain isolated. Although it is evident at the bottom of each network snapshot that the network includes a large pool of non-connected co-authorship groups, the ETS Journal co-authorship network tends toward developing a focused topology.
Conclusions and future work

Co-authorship networks analysis can provide interesting insights about how research communities are collaborating and gradually evolve. Nevertheless, there are no co-authorship network analysis studies for networks formed by the authors of scientific journals in the field of TeL. Thus, in this paper we study the co-authorship network formed by the authors of the ETS Journal, an open access accredited academic journal dedicated to TeL research, where 792 co-authored papers were analyzed. Within this network, we have identified popular, central and influential authors, major clusters of authors collaborating on specific research areas and interesting characteristics about the network evolution. More precisely, it has been identified that:

- Key authors (namely, those with high centrality metrics) of the network are coming from Taiwan. Taiwanese researchers have established a strongly connected group that collaborates frequently, diversely and widely. Moreover, we have collected evidence for the correlation between the number of citations and three centrality metrics, namely degree centrality, betweenness centrality and closeness centrality.

- The network includes different subgroups of authors that collaborate independently. However, the network has probably arrived at its “phase transition” in 2010 and it is expected that the authors will start collaborating more frequently and more widely in the coming years. This means that the network is expected to become more connected and its diameter will be decreased.

Drawing from the analyses in this paper, interesting recommendations can be extracted as follows:

- For the ETS journal editors, their efforts in retaining the key authors (namely, those with high centrality metrics) of the journal active are very important for the future evolution of the co-authorship network. More specifically, ETS editors can approach key authors for organizing special issues in specific research topics or they can offer them roles in the editorial board of the journal.

- For the key authors of the network, they should be active in finding, suggesting and setting up new collaborations with members in different sub-groups, particularly from the journal network’s periphery, which will make the entire network more integrated and connected. Key members also play an important role in engaging new authors and connecting them to the core of the co-authorship network. This can also lead to the introduction of emerging research ideas or topics that come from these new authors to the current main research areas.

- For the potential authors/TeL researchers, they should take into consideration the current structure of the ETS co-authorship network, before submitting a paper for publication, since they can benefit from the knowledge of research collaboration patterns and key research areas of collaboration.

In future work we are planning to augment the findings drawn from this study by considering the data from the latest issues of ETS journal (namely, issues published in 2013 and 2014) as well as data from other journals in the field of TeL. This will enable us to generalize our findings and compare them to other research fields towards identifying similarities or differences, as well as to have a complete view of the research collaboration that takes place in the TeL landscape. Moreover, another dimension for our future work will be the spatial proximity factor of research collaboration, which can be studied by identifying collaboration patterns between authors from different countries. Finally, our future efforts will also focus on building research analytics tools and/or services that will be able to visualize and further analyze the data that have been extracted from this study.

Acknowledgements

The work of Demetrios G. Sampson has been partially supported by the International Research Collaboration project funded by the National Science Council under grant number: NSC101-2911-I-110-508

References


A Review of Research on Mobile Learning in Teacher Education

Evrim Baran
Middle East Technical University, Faculty of Education, Department of Educational Sciences Ankara, Turkey // ebaran@metu.edu.tr

ABSTRACT
Mobile devices have become attractive learning devices for education. While the majority of the existing research has focused primarily on the value of mobile learning for students, researchers have recently started exploring its potentials within teacher development. The present qualitative synthesis of quantitative and qualitative research aimed to address trends and gaps observed in the literature regarding the integration of mobile learning into teacher education. Six main findings emerged: (a) there is an increasing trend in integrating mobile learning in teacher education contexts; (b) theoretical and conceptual perspectives are scarcely reported; (c) variations exist in perceptions, attitudes and usage patterns; (d) engagement with mobile learning and devices is primarily reported as being beneficial; (e) challenges were scarcely reported; and (f) several pedagogical affordances support mobile learning integration into teacher education settings. These findings have been interpreted to determine their implications on the development of mobile learning experiences in teacher education, including programmatic directions for integration and study.

Keywords
Mobile learning, Teacher education, M-learning, Preservice, In-service

Introduction
Over the past two decades, technology devices have become mobile — portable and networked — to the point that they have become pervasive in everyday life. The use of mobile devices has become common among a wide range of age groups due to affordability and availability (Newhouse, Williams, & Pearson, 2006). Significant investments have been made to provide infrastructure, content, and resources related to the integration of mobile devices into learning environments (Johnson, Smith, Willis, Levine, & Haywood, 2011), and researchers have long had an interest in this evolving landscape (Kukulska-Hulme, Sharples, Milrad, Arnedillo-Sánchez, & Vavoula, 2009). However several limitations exist, such as lack of theoretical and pedagogical underpinnings, sustainable integration into formal educational contexts, and, particularly, lack of teacher support and training (Cochrane, 2012; Peng, Su, Chou, & Tsai, 2009).

Teacher support and teacher training have been the least explored topics in mobile learning research (Ekanayake & Wishart, 2014). Mobile learning is especially under-theorized in teacher education (Kearney & Maher, 2013), despite the need to inform teachers of the value of mobile technologies and how to integrate them effectively into their classes (Schuck, Aubusson, Kearney, & Burden, 2013). In their review of mobile learning projects conducted in Europe, Kukulska-Hulme et al. (2009) revealed that at the “European and individual state level, there appears to be little teacher development or training activity addressing mobile learning” (p. 14). Challenges related to teachers’ adoption of mobile technologies have emerged from the fact that they are not effectively prepared to investigate the advantages or make informed decisions (Kukulska-Hulme et al., 2009; Schuck et al., 2013). Because of both the pressure to provide teachers with effective technology integration skills and the rapid growth of mobile technologies as learning devices, teacher education programs need to implement theoretically and pedagogically sound mobile learning initiatives (Newhouse et al., 2006).

This review aims to fill a gap in the current research on mobile learning. Previous literature reviews have synthesized trends and provided analysis of findings (Hwang & Tsai, 2011; Hung & Zhang, 2012), but no systematic research has been conducted on mobile learning and teacher education. This is the first review to initiate an evidence-based discussion on mobile learning and related emerging pedagogical directions in teacher education.

Mobile learning and theoretical perspectives
The diversity of the research on mobile learning has made it difficult to generate a single definition or to determine generally added benefits (Frohberg, Göth, & Schwabe, 2009; Sharples, Arnedillo-Sánchez, Milrad, & Vavoula, 2009). While it is typical for an emerging field to have varied definitions, the lack of conceptual frameworks and
robust theories has been frequently addressed as a concern in the literature (Peng et al., 2009). Definitions of mobile learning emphasize mobility (Sharples et al., 2009), access (Parsons & Ryu, 2006), immediacy (Kynäslahti, 2003), situativity (Cheon, Lee, Crooks, & Song, 2012), ubiquity (Kukulska-Hulme et al., 2009), convenience (Kynäslahti, 2003), and contextuality (Kearney, Schuck, Burden, & Aubusson, 2012). According to Sharples et al. (2009), mobile learning includes the characteristics of mobility in physical, conceptual, and social spaces. The “relationship between the context of learning and context of being” is unique to mobile learning, as learning may occur in independent, formal, or socialized contexts (Frohberg et al., 2009, p. 313).

The greatest added value of mobile learning vis-a-vis PC learning lies in the aspects that extend classroom interaction to other locations via communication networks. Recent advances such as imbedded sensors, cameras, motion detection, location awareness, social networks, web searching, and augmented reality present the potential to foster learning and engagement across multiple physical, conceptual, and social spaces, both indoors and out (Newhouse et al., 2006). Mobile learning enables teachers and learners ubiquitous and seamless access to information (Kukulska-Hulme et al., 2009; Seppälä & Alamäki, 2003), and convenience, expediency, and immediacy are valuable to teachers and enhance students’ learning (Kynäslahti, 2003). These features provide opportunities for individualized, situated, collaborative, and informal learning without being limited to classroom contexts (Cheon et al., 2012). While portability and mobility have already made these devices attractive tools, developments such as geospatial technologies, search capabilities, image and video capture, and context awareness have further increased their versatility by promoting situated learning experiences and allowing exploration within authentic settings, particularly supporting inquiry-based learning (Martin & Ertzberger, 2013).

While the majority of research on mobile learning has focused primarily on students, recently teachers and researchers have started exploring the potentials of mobile learning and devices within teacher education. By synthesizing the literature on mobile learning and teacher education, this research aimed to address the trends and gaps observed in order to determine current implementation practices.

**Research methods**

Educational research syntheses demonstrate important interactions and connections from existing literature and offer conclusions or build theories for further research and practice (Minner, Levy, & Century, 2010). Given the methodological diversity of educational research on mobile learning in teacher education, this review is a qualitative synthesis of quantitative and qualitative research. Qualitative research syntheses can be defined as, “[S]ystematic efforts of synthesizing qualitative research” (Suri & Clarke, 2009, p. 401). This research synthesis departed from a strictly positivist, meta-analytic approach to syntheses, with its promotion of, “[T]ransparency of process to enhance accountability, credibility, and transferability of synthesis findings” (Suri & Clarke, 2009, p. 413) as suggested for research syntheses primarily qualitative in nature.

The synthesis followed three distinct phases: search and inclusion, individual study review, and cross-study comparison and analysis. Phase 1 involved searching for keywords “mobile learning” and “teacher education” or “mobile learning” and “teacher” or “mobile” and “teacher” in academic journals in the ERIC and Education Research Complete databases, and this search yielded 92 results. In the second round of search, the search terms “teacher education” and “mobile learning” were applied to Google Scholar and the following primary research journals: *Journal of Educational Technology and Society, Computers and Education, British Journal of Educational Technology, Journal of Computer Assisted Learning, Educational Technology Research and Development, Journal of Computing in Teacher Education, Journal of Technology and Teacher Education, Journal of Research on Technology in Education, Journal of Digital Learning in Teacher Education*. Moreover, references in each of the identified articles were checked for related work. As of August 2014, 237 new articles were thus identified from this search, resulting in a new total of 329 articles. The articles were organized and coded according to their context (e.g., preservice teacher education, in-service teacher education, professional development, K12 classrooms), and type (e.g., empirical, theoretical, case description, editorial, policy).

For the final tally of articles for the research synthesis, inclusion criteria were applied to ensure the articles included: (a) empirical research on mobile learning in preservice and in-service teacher education contexts across different disciplines (e.g., social studies, literacy, and math); (b) applications of mobile technologies (e.g., mobile phones,
smartphones, tablets) in a teacher education context; (c) in-service teacher, preservice teacher, or teacher educator participants; and (d) publication in a peer-reviewed journal, rather than a technical reports, project anecdote, or similar proceeding. Studies on preservice and in-service teachers’ perceptions of mobile learning were also included to present the current landscape surrounding the use of these technologies as teaching tool. Because the aim was to investigate empirical research on mobile learning in teacher education contexts, these exclusion criteria significantly decreased the number of articles to 42. 5 more articles were also eliminated because they presented a theoretical or conceptual work on mobile learning and not empirical research. Two articles that lacked the descriptions of their research methodologies were included in the list because they provided anecdotal evidence regarding mobile learning and its implications in teacher education and helped to broaden the focus of this review.

Phase 2 included an analysis of the remaining 37 articles. The articles were subsequently examined and coded with notations in an analytic research synthesis table including the meta-categories of teacher education context, country context, subject domain, type, definition of mobile learning, reliability, validity and trustworthiness reports, mobile technologies used, outcomes in terms of teacher knowledge and practice, and pedagogical approaches. Notations related to methodology, such as the study’s purpose, data sources, and participant information were included in the analytic table to foster comparisons among the studies (see Table 1). The salient aspects of the articles and their categories were evaluated with two research assistants as external reviewers who had research experiences in mobile learning and teacher education projects.

Phase 3 of the research included a comparison of the studies within the identified categories. The data were further investigated “to consider themes, shapes, and organization of research ideas present in the overall literature” (Opfer & Pedder, 2011, p. 38). Following themes emerged that captured variations and commonalities: Trends, perceptions, attitudes and usage patterns, benefits and challenges of mobile learning, and practices within teacher education contexts. The findings below report on these comparisons with conclusions representing the research synthesis and implications.

Table 1. An analysis of studies on mobile learning and teacher education (n = 37)

<table>
<thead>
<tr>
<th>Study</th>
<th>Subject Domain</th>
<th>Subjects</th>
<th>Type</th>
<th>Method</th>
<th>Data sources</th>
<th>Reliability, validity, trustworthiness report</th>
<th>Technology used</th>
<th>Country context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ekanayake &amp; Wishart (2014)</td>
<td>Science</td>
<td>In-service teachers</td>
<td>Implementation</td>
<td>Case study</td>
<td>Observations via video, audio recording and field notes</td>
<td>* Mobile phone</td>
<td>Sri Lanka</td>
<td></td>
</tr>
<tr>
<td>Husbye &amp; Elsener (2013)</td>
<td>Literacy</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Case description</td>
<td>--</td>
<td>Tablet, smartphone, laptop</td>
<td>U.S.A.</td>
<td></td>
</tr>
<tr>
<td>Foulger et al. (2013)</td>
<td>Mix</td>
<td>Teacher educators</td>
<td>Survey</td>
<td>Interpretivist model of qualitative research</td>
<td>Questionnaire</td>
<td>* Mobile device</td>
<td>U.S.A.</td>
<td></td>
</tr>
<tr>
<td>Aubusson, Schuck, &amp; Burden (2009)</td>
<td>Mix</td>
<td>In-service teachers, teacher advisors and developers</td>
<td>Survey</td>
<td>Qualitative</td>
<td>Interview</td>
<td>Mobile phone</td>
<td>Cross country (Australia, UK)</td>
<td></td>
</tr>
<tr>
<td>Schuck et al. (2013)</td>
<td>Mix</td>
<td>Teacher educators</td>
<td>Implementation</td>
<td>Design-based research</td>
<td>Written materials, artifacts, discussions</td>
<td>* Smartphone, iPod</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>Serin (2012)</td>
<td>Mix</td>
<td>Preservice teachers</td>
<td>Survey</td>
<td>Mixed methods</td>
<td>Questionnaire</td>
<td>* Mobile technology</td>
<td>Turkish Republic of Northern Cyprus</td>
<td></td>
</tr>
<tr>
<td>Uzunboylu &amp; Özdamlı (2011)</td>
<td>Mix</td>
<td>In-service teachers</td>
<td>Survey</td>
<td>Survey development</td>
<td>Questionnaire</td>
<td>* M-learning technology</td>
<td>Turkish Republic of Northern Cyprus</td>
<td></td>
</tr>
<tr>
<td>Seppälä &amp; Alamäki (2003)</td>
<td>Home Economics</td>
<td>Preservice teachers, supervising</td>
<td>Implementation</td>
<td>Case study</td>
<td>Interview</td>
<td>Smartphone</td>
<td>Finland</td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Area</td>
<td>Research Design</td>
<td>Methods</td>
<td>Tech Used</td>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-----------------</td>
<td>---------</td>
<td>-----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cushing (2011)</td>
<td>Preservice teachers, supervising teachers, project leaders, support staff</td>
<td>Implementation</td>
<td>Case study</td>
<td>Interviews, observations, retention and achievement data, survey</td>
<td>Smartphone</td>
<td>UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thomas, O’Bannon, &amp; Bolton (2011)</td>
<td>In-service teachers</td>
<td>Survey</td>
<td>Survey</td>
<td>Survey</td>
<td>* Mobile phone</td>
<td>U.S.A.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Şad &amp; Göktas (2013)</td>
<td>Preservice teachers</td>
<td>Survey</td>
<td>Survey</td>
<td>Survey</td>
<td>* Mobile phone, laptop</td>
<td>Turkey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kearney &amp; Maher (2013)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Case Study</td>
<td>Journals, focus group interviews, artifacts</td>
<td>iPad</td>
<td>Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newhouse et al. (2006)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Mixed methods</td>
<td>Survey</td>
<td>Laptop</td>
<td>Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen (2010)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Mixed methods</td>
<td>Survey</td>
<td>PDA</td>
<td>Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahruf, Shohel, &amp; Power (2010)</td>
<td>In-service teachers</td>
<td>Implementation</td>
<td>Mixed methods</td>
<td>Survey, classroom observation, interviews</td>
<td>iPod</td>
<td>Bangladesh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hashim (2014)</td>
<td>In-service teachers</td>
<td>Implementation</td>
<td>Quantitative</td>
<td>Survey</td>
<td>* iPad</td>
<td>Malaysia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ismail, Azizan, &amp; Azman (2013)</td>
<td>Information Technology</td>
<td>Survey</td>
<td>Quantitative</td>
<td>Questionnaire</td>
<td>* Mobile phone</td>
<td>Malaysia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kommers (2009)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Experimental</td>
<td>Survey</td>
<td>PDA</td>
<td>The Netherlands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bates &amp; Martin (2013)</td>
<td>Literacy coaches</td>
<td>Implementation</td>
<td>Design based research</td>
<td>Interviews, blog, digital notes</td>
<td>iPad</td>
<td>U.S.A.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valtonen, Havu-Nuutinen, Dillon, &amp; Vesisenaho (2011)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Case study</td>
<td>Students’ lecture notes</td>
<td>Laptop</td>
<td>Finland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciampa (2014)</td>
<td>In-service teachers</td>
<td>Survey</td>
<td>Case study</td>
<td>Interview, teacher blog, observation, ecological surveys</td>
<td>Tablet</td>
<td>Canada</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burton et al. (2011)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Mixed methods</td>
<td>Pre-post test, cell phone attitude survey</td>
<td>* Mobile phone</td>
<td>U.S.A.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franklin, Sexton, Lu, &amp; Ma (2007)</td>
<td>Preservice teachers</td>
<td>Implementation</td>
<td>Qualitative Case study</td>
<td>Weekly journals, pre-post survey of technology skills, classroom observation,</td>
<td>PDA</td>
<td>U.S.A.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Several main findings emerged as a result of the research synthesis of the selected 37 articles on mobile learning and teacher education, outlined below in terms of trends, theoretical perspectives, perceptions, attitudes, and usage patterns, benefits, challenges, and teacher education contexts.

**Trends in the mobile learning and teacher education literature**

The articles examined were bound within the period, between 2000 and the first half of 2014, when research on emerging mobile technologies in education was burgeoning. As illustrated in Figure 1, in the last two years more than 50% of the articles from the period 2000-2014 have been published. Consistent with the widespread usage of
mobile tools, researchers and teacher educators have showed an increasing interest in the integration of mobile technologies into teacher education contexts. As shown in Table 1, almost 38% of the research on mobile learning was conducted in the U.S. A teacher education contexts followed by Australia (n = 3) and Finland (n = 3). It is especially noteworthy that more countries contributed to the mobile learning research in teacher education in recent years, such as Malaysia, Canada, Singapore, UK and Tanzania.

![Figure 1. Publication trends from 2000 to 2014](image)

The studies examined in this review used varied methodologies, with a majority being case studies or mixed method. Of the 37 studies, 21 were conducted within preservice teacher education contexts, 10 within in-service teacher education context, four within teacher educators’ contexts, one in both preservice and in-service teacher education contexts, and one in a coaching context. As shown in Figure 2, the majority of the studies crossed content areas, except five studies on science, three on math, three on literacy, two on physical education, one on information technology, one on STEM teacher education, one on English language teaching, and one on home economics teacher education.

![Figure 2. Distribution of studies by subject domain](image)

The articles’ research purposes were classified according to two types: (1) survey-dominant or (2) implementation-dominant. More than half of the studies were identified as implementation-dominant (n = 25) that focused on the application of a mobile learning system, program, project or course and presentation of cases and strategies of mobile learning in teacher education contexts, while others were identified as survey-dominant (n = 12) that mainly focused
on surveying participants’ usage, perceptions, or attitudes on mobile tools within their contexts. Data sources included questionnaires, interviews, blogs, recordings, observations, journals, artifacts, usage data, and audio and video transcripts. Among 37 studies reviewed, only 17 articles reported the validity, reliability or trustworthiness of their measures. Mobile phones (e.g. smartphones, cell phones) were the most common mobile devices used in teacher education contexts (42.5%), followed by tablets (e.g., iPads) (17.5%), PDAs/Handheld PCs (17.5%), iPods (10%), and laptops (12.5%). The type of mobile device used over time, particularly in 2013 and 2014, mostly included mobile phones/smartphones and tablets. The large number of implementation studies illustrates teacher educators’ interest in the application of mobile learning systems in their contexts. With mobile technologies being more ubiquitous, it is predicted that teacher educators will continue to explore the pedagogical affordances of new mobile technologies in their contexts.

**Scant report of theoretical and conceptual perspectives**

Analysis revealed scant reports of theoretical perspectives integrated into mobile learning research in teacher education. Out of 37 studies, only five reported using a theoretical or pedagogical framework to design or implement the research. Kearney et al. (2012), using m-learning and socio-cultural theory, identified three distinctive features of m-learning: authenticity, collaboration, and personalization. These features were later tested within m-learning projects conducted in teacher education communities (Kearney & Maher, 2013). Kearney and Maher (2013) emphasized the importance of putting pedagogy at the center of mobile learning rather than technology in order to examine its advantages for supporting learning. In order to help teacher educators understand mobile technology integration, Schuck et al. (2013) combined the concept of a professional learning community (PLC) with a community of practice (CoP) and coined the term mobagogy to “capture dual interests of the community in mobile technologies and pedagogy” (p. 4). The purpose of this approach was to help teacher educators explore mobile learning through their interactions with colleagues in the CoP. Other theoretical perspectives included experiential self-regulated learning (Järvelä et al., 2007), motivation theory (Ciampa, 2014), and cognitive development theory (McCaughtry & Dillon, 2008). Järvelä et al. (2007) used self-regulated learning theory “as a theoretical framework to develop those learning activities that give potential to individual and collaborative learning so that it stimulates active minds and interactions on individual and social levels” (p. 72). Ciampa (2014) used motivation theory to understand how motivational elements could help design of learning systems with mobile tools, and McCaughtry and Dillon (2008) used cognitive development theory to examine changes in teachers’ thinking as they integrate new technologies in their contexts.

The literature investigated herein lacks new approaches, models and frameworks for teacher education programs designed specifically to develop mobile learning pedagogies. Mobile tools have the potential to enhance mobility in classrooms; fundamentally changing the way classrooms are organized within teacher education programs. However, the literature needs to establish pedagogical and theoretical models that can guide teacher educators in designing mobile learning experiences for preservice and in-service teachers. These models need to present strategies for equipping teachers and teacher educators with methods for integrating mobile learning into classrooms as well as supporting professional learning with mobile tools. Similarly, systematic and programmatic efforts are missing that explore the integration of mobile learning into preservice teacher education curricula.

**Varied perceptions, attitudes and usage patterns**

Survey studies conducted with preservice teachers revealed varied results regarding perceptions, attitudes and usage about the use of mobile devices for learning and teaching. The outcomes were diverse depending on the variables in teacher education contexts such as the availability and accessibility of technologies, resources, country infrastructure, and motivation of teacher education programs. Researchers found that an increasing number of preservice teachers were accessing resources on mobile devices (Hossain & Quinn, 2013; Shotsberger, 2003). However, Serin’s (2012) investigation of 355 prospective teachers revealed low levels of perceptions regarding mobile learning. Similarly, Thomas and O’Bannon’s (2013) survey study with preservice teachers revealed that perceived benefits of cell phone usage in the classrooms were limited. Şad and Göktaş’s (2013) survey study reported that preservice teachers favored laptops over mobile phones lacking smartphone features. McCaughtry & Dillon (2008) reported a shift in preservice teachers’ perceptions following initial skepticism towards integration of mobile devices (PALMS) in teacher education programs.
education, and Gado et al. (2006) reported a positive change in preservice teachers’ self-efficacy and attitudes towards technologies after classroom exposure to handheld computers.

Survey research conducted with in-service teachers also reported varied results. For example, Uzunboylu and Özdamlı’s (2011) investigation of the m-learning perception scale with 467 teachers in the Turkish Republic of Northern Cyprus indicated above medium levels of perception regarding mobile learning. However, another survey conducted with 38 Malaysian in-service teachers revealed that the majority did not consider mobile phones learning and teaching tools within their schools. On the other hand, Thomas, O’Bannon and Bolton’s (2013) survey study of 79 teachers found that the majority supported the use of cell phones in the classroom. O’Bannon & Thomas’s (2014) survey on the use of mobile phones in the classroom with 1095 teachers in the southeastern United States revealed that teachers older than 50 were less supportive compared to younger groups.

This review revealed scant report of validity and reliability issues in the survey studies as well as information on piloting the instruments. Only seven studies provided some explanation of a pilot test conducted on their data collection instruments (e.g., Hashim, 2014; Ismail, Azizan, & Azman, 2013; O’Bannon & Thomas, 2014; Şad & Göktaş, 2013; Thomas & O’Bannon, 2013; Thomas, O’Bannon, & Bolton, 2013; Uzunboylu & Özdamlı, 2011). Limited information on these issues makes it difficult to compare the results of these questionnaires. Research in this area would further benefit from discussion on the trustworthiness of the methods and more detailed information about the context used to investigate mobile learning in teacher education.

**Mobile learning is reported as mainly beneficial**

The literature on mobile learning and teacher education generally considered mobile learning a beneficial approach in extending teachers’ learning experiences and enhancing their mobile technology integration skills. The literature identified several motivating factors for the integration of mobile learning into preservice teacher education settings, such as modeling mobile pedagogies (Burton et al., 2011; Cushing, 2011; Foulger et al., 2013; Franklin et al., 2007; Gado et al., 2006; Herro et al., 2013; McCaughtry & Dillon, 2008), deeper explorations of content areas (Gado et al., 2006; Husbye & Elsener, 2013; Kearney & Maher, 2013; Mahruf et al., 2010; McCaughtry & Dillon, 2008; Shotsberger, 2003; Burton et al., 2011), enhancing preservice teachers’ mobility (Coens et al., 2011; Husbye & Elsener, 2013), connecting preservice teachers with a larger community (Cushing, 2011; Husbye & Elsener, 2013; Kearney & Maher, 2013), providing preservice teachers with personalized learning experiences (Kommers, 2009), enhancing social interaction (Järvelä et al., 2007; McCaughtry & Dillon, 2008; Valtonen et al., 2011), presenting alternative assessment techniques (Chen, 2010), and promoting collaborative knowledge construction (Järvelä et al., 2007).

Studies that investigated the use of mobile learning or mobile tools in teacher education contexts mainly reported positive contributions to the outcomes investigated. Mobile tools were found to have potential for helping preservice teachers understand and develop new literacies (Husbye & Elsener, 2013), explore mathematics in the real world (Kearney & Maher, 2013; Shotsberger, 2003), conduct scientific investigations (Gado et al., 2006), engage in rich language learning contexts (Mahruf et al., 2010); and explore real world physical education (McCaughtry & Dillon, 2008). These tools can fundamentally change the way classrooms are organized within teacher education programs by increasing mobility (Price et al., 2014). Another advantage reported was features that connect preservice teachers to their colleagues, enhance professional learning through collaboration, and facilitate mentoring processes (Cushing, 2011; Husbye & Elsener, 2013; Kearney & Maher, 2013). Mobile tools could help build closer relationships as well as more personalized learning experiences for teacher candidates as needs change over time (Crippen & Brooks, 2000; Herro et al., 2013; Kommers, 2009).

Advantages in integrating mobile learning into in-service teacher education contexts included promoting reflection-in-action as a critical component of professional learning (Aubusson et al., 2009); providing timely access to resources (Shotsberger, 2003); allowing participation in knowledge production and sharing regarding teaching practices (Aubusson et al., 2009); and capturing, reflecting upon, and sharing experiences (Aubusson et al., 2009). In addition to conducting professional development via mobile tools, researchers also investigated the impact of such programs on teachers’ inclusion of these technologies in their classrooms. Looi et al. (2014) investigated a mobilized 5E science curriculum co-designed by the teachers and observed that their pedagogical orientations affected both their technology integration and their relationships with students.
Advantages of mobile learning for teacher educators were also addressed in the literature. Husbye & Elsener (2013) found that after being exposed to teacher educators’ mobile device integration, preservice teachers began to utilize such tools in their own practices. By encouraging mobility, the teacher educators’ role shifted from a content provider at the center of instruction to a facilitator, engaging preservice teachers as they collaboratively constructed meanings around content (Husbye & Elsener, 2013).

Scant report of challenges

This research synthesis revealed that extant studies have mainly emphasized the benefits of mobile learning integration into teacher education without detailing its drawbacks. According to the results of Foulger et al.’s (2013) survey in the United States, teacher educators are in the process of investigating this innovation further, as “mobile technology in teacher preparation is uncharted territory and they are taking a certain level of risk by exploring its possibilities” (Foulger et al., 2013, p. 22). A number of challenges related to mobile technology integration were reported, including ethical issues, lack of support, accessibility and technical limitations, insufficient experience, mobile phone bans in schools, and curriculum adaptations.

Teachers, while accepting the benefits of mobile learning, raised concerns about potential ethical issues such as cyber-bullying, privacy, archiving and record keeping, sharing classroom experiences and artifacts, parental and student informed consent, and e-safety (Aubusson et al., 2009; Cushing, 2011). Teacher educators, by encouraging ethical use of these devices as learning tools to enhance rather than disrupt class flow, could model practices for preservice and in-service teachers (Cushing, 2011).

Another challenge noted was the need for ongoing technical and material support (Mahruf et al., 2010); teachers received minimal technological and pedagogical assistance from higher education institutions with regard to effective implementation of mobile learning in teacher education (Foulger et al., 2013; Cushing, 2011). Lack of support in terms of policies could also lower teachers’ perceptions regarding the use of mobile devices as learning tools (Ismail et al., 2013), as well as insufficient funding or professional development support (Herro et al., 2013).

The accessibility of mobile devices is another challenge. If mobile learning is to be implemented successfully, all preservice and in-service teachers must have access to mobile devices as part of their training (Cushing, 2011; Gado et al., 2006; McCAU ghy & Dillon, 2008). Teacher educators noted that mobile devices should be provided to preservice teachers to ensure digital equity (Husbye & Elsener, 2013). Additional technical limitations included low bandwidth on wireless networks, small screen size, insufficient memory capacities, and limited software (Franklin et al., 2007; Newhouse et al., 2006).

Lack of expertise integrating mobile technologies was also a challenge to effective integration of mobile learning into teacher education (Foulger et al., 2013; Valtonen et al., 2011). Research has suggested that mobile technologies need to be used as tools for enhancing preservice teachers’ experiences, not as an add-on incorporating technology for its own sake (Husbye & Elsener, 2013). Mobile learning must be meaningfully integrated into all teacher education courses, not only technology courses (Foulger et al., 2013). Limited, unclear best practices regarding preparing teachers for the integration of mobile devices is a definite barrier (Herro et al., 2013).

One particularly critical challenge noted with regard to integration of mobile learning into in-service teacher education is the prohibition of cell phone usage within schools, which may affect teachers’ attitudes regarding mobile learning in their classes or prevent them from making any efforts to that end (Ismail et al., 2013).

If mobile tools are to be integrated effectively into classrooms, curricular issues also need to be taken into consideration. For example, Looi et al. (2014) found limited research on the investigation of teachers’ curricular-based implementations of mobile learning and devices. Price et al. (2014) further noted that while preservice teachers’ ideas about integration into science classes could be supported by teacher education programs, their implementation into a non-existing or incompatible curriculum (e.g., geospatial integration in science) is obviously an obstacle.

While benefits were commonly reported in the literature, the studies lacked investigations on the drawbacks and challenges of mobile learning in teacher education. For example, the discussions on the issues regarding ethical use,
accessibility, and privacy were limited. It is thus crucial to conduct longitudinal research to determine the impact of sustained mobile learning initiatives in teacher education programs and critical success factors as well as challenges.

Mobile learning practices within teacher education contexts

The analysis of mobile learning practices within the studies revealed approaches with different goals based on audience. Three groups participated in the studies: preservice teachers, in-service teachers, and teacher educators.

*Integrating mobile learning into preservice teacher education*

Preservice teacher education programs and courses were the most common research contexts in the literature (e.g., Husbye & Elsener, 2013). Additionally, studies were conducted in the context of projects that included preservice teachers as participants of the studies (e.g., Schuck et al., 2013) and postgraduate teacher education certification programs (e.g., Price et al., 2014). The research synthesis revealed many pedagogical advantages of mobile learning in preservice teacher education: connectivity and collaboration, flipped classroom models, mobility within the physical space of the classroom, backchannel conversations, engaging with content on mobile devices, mobile learning in student teaching, performance evaluation, and participation in PLCs.

Mobile tools have capabilities of capturing real time information and integrating connectivity and collaboration into class activities. For example, Husbye and Elsener (2013) integrated mobile phones into early childhood literacy courses using QR codes to index videos of children reading to help preservice teachers move at their own pace. Mobile devices’ connection capabilities also provided opportunities to share preservice teachers’ products (e.g., teaching videos) on the web (Husbye & Elsener, 2013; Schuck et al., 2013). Similarly, Järvelä et al. (2007) used a mobile mind map tool to co-regulate preservice teachers’ collaborative knowledge construction, as mobile devices “supported externalization of knowledge representations in individual and collaborative levels” (p. 71).

Flipped classrooms have recently been used as form of blended learning in which students learn course content on the web via video, audio, or text and use class time to engage in activities and get individual guidance. In teacher education courses, mobile devices helped establish flipped classrooms. For example, Husbye and Elsener (2013) asked preservice teachers to access materials (e.g., video podcasts) before class and engage in hands-on activity during class.

Mobile devices allow for mobility within and beyond the classroom’s physical spaces. Another example shared by Husbye and Elsener (2013) was digital gallery walks, where students accessed web resources linked to QR codes on posters in class. Other examples included podcasts in contexts such as botanical gardens (Schuck et al., 2013) and capturing real life math phenomena (Kearney & Maher, 2013). Burton et al. (2011) implemented an augmented reality (AR) mobile game to enhance preservice teachers’ self-efficacy and attitudes towards integrating AR pedagogies in future STEM classrooms. Following the participatory design approach, Price et al. (2014) designed the GeoSciTeach smartphone application to support preservice science teachers’ awareness of the integration of geospatial ideas into science. Mobility, combined with other emerging features such as augmented reality and context awareness, helped facilitate contextualized and situated learning experiences.

The timely communication afforded by mobile devices could transform a class conversation from one-to-many to many-to-many and increase students’ engagement via backchannel conversations. Teacher educators integrated mobile devices into courses where preservice teachers shared their understanding of content and participated in conversation online. Husbye and Elsener (2013), for example, used class-specific hashtags on Twitter, where students discussed class activities, commented on classroom experiences, and shared resources for best teaching practices. Similarly, Valtonen et al. (2011) used mobile devices in teacher education courses to enable students to capture and share lecture notes via social software, while Järvelä et al. (2007) used a mobile lecture interaction tool to enhance participation during lectures. Finally, Schuck et al. (2013) conducted activities where preservice teachers voted through text messages or “quick response” voting applications.

Preservice teachers also engaged with content on mobile devices. One group created digital narratives by capturing and editing videos, sharing them at a mobile phone film festival (Schuck et al., 2013). A science methods course
included the use of handhelds to engage preservice teachers in scientific inquiry to develop their understanding of science and math concepts (Gado et al., 2006). Preservice teachers also used mobile devices to organize their work and access reference tools such as dictionaries or the periodic table (Franklin et al., 2007).

Due to the potential for enhancement, researchers have investigated the integration of mobile devices into student teaching experiences. Preservice teachers, mentors, and teacher educators can easily connect via mobile tools to share feedback (Crippen & Brooks, 2000; Cushing, 2011; Foulger et al., 2013; Kommers, 2009; Schuck et al., 2013; Seppälä & Alamäki, 2003). Examples include microblogging (Schuck et al., 2013); virtual training (Seppälä & Alamäki, 2003); connecting mentors, project leaders, and preservice teachers via smartphones (Cushing, 2011) and PDAs (McCaughrty & Dillon, 2008); submitting school observation forms and weekly e-journals (Crippen & Brooks, 2000; Shotsberger, 2003); designing lesson plans with mobile device integration (Foulger et al., 2013); and real-time coaching (Kommers, 2009).

Finally, mobile devices allow preservice and in-service teachers to evaluate their own and their peers’ learning quickly and efficiently. Chen (2010) developed a Mobile Assessment Participation System (MAPS) that aimed to facilitate the assessment of preservice teachers’ self, peer, and group performances both synchronously and asynchronously. Preservice teachers also participated in PLCs on mobile devices, where they expanded on conversations and improved their understandings (Schuck et al., 2013). Mobile devices encouraged dialogues and communication between preservice teachers and practicum supervisors as well as teacher educators (Seppälä & Alamäki, 2003) and were used for synchronous coaching in school-based practicums, where preservice teachers received immediate, “on the job” feedback (Kommers, 2009).

**Mobile learning with in-service teachers**

Among 11 studies that investigated in-service teachers, only four of them implemented professional development programs on mobile learning, while others focused on surveying teachers’ perceptions and attitudes towards mobile devices and mobile learning. Mobile learning research conducted within in-service teacher education contexts focused on their experiences using mobile technologies for their own learning (Aubusson et al., 2009; Hashim, 2014), the impact of mobile learning professional development programs on effective integration of mobile devices into classrooms (Ekanayake & Wishart, 2014; Mahruf et al., 2010), and perceptions regarding the use of mobile devices (Aubusson et al., 2009; Ciampa, 2014; Uzunboylu & Özdamlı, 2011).

Mobile learning holds promises for creating mobile, collaborative, contextualized, customized, and personalized learning opportunities for teachers. Aubusson et al.’s (2009) interviews with eight teachers, teacher developers, and teacher advisors in Australia and the UK revealed that educators considered mobile technology as beneficial for enhancing on the job learning. Uzunboylu and Özdamı’s (2011) mobile learning perception scale conducted revealed in-service teachers’ (n = 467) positive perceptions toward m-learning. Another survey with in-service teachers revealed that iPads helped them access learning materials, collaborate in online forums, and access email. Ciampa’s (2014) case study of teachers’ perceptions on how mobile devices motivated students revealed the “six key aspects of successful (mobile) learning systems as challenge, control, curiosity, recognition, cooperation and competition” (p. 92).

To equip in-service teachers with the skills for effective integration of mobile devices, researchers conducted professional development programs and assessed their effectiveness. Such programs included the design of a series of workshops for secondary science teachers (Ekanayake & Wishart, 2014) and mobile access to development resources for English teachers (Mahruf et al., 2010). Ekanayake and Wishart’s (2014) analysis of paper-based materials and audio and video recordings revealed that their workshops increased awareness of and response toward the potential benefits of science teaching and learning with mobile phones, helped develop pedagogical actions, and encouraged sharing knowledge related to technology integration. In another project, teachers watched and listened to material on iPods and other mobile devices (Mahruf et al., 2010). Interviews with participants revealed that accessing professional development resources via mobile devices helped them engage in new settings and develop pedagogical knowledge and English-language proficiency.

With the recent emphasis on using mobile technologies in education, the need has emerged to prepare teachers with effective classroom technology integration skills. Looi et al.’s (2014) unique study included teachers in the
curriculum design team and observed their mobile lesson enactment in classroom contexts. Similar programs for teacher development that follow situated, reflective, and authentic models hold promises for supporting teachers' learning of mobile technology integration.

**Mobile learning with teacher educators**

With the widespread use of mobile devices, teacher educators are expected to show more interest in testing the strategies for implementing mobile devices in their courses and professional development programs. Methods for preparing teacher educators for mobile learning included PLCs (Schuck et al., 2013) and literacy coaching practices (Bates & Martin, 2013). Using design-based research, Schuck et al. (2013) employed a PLC to support teacher educators in understanding advantages of mobile learning within teacher education and model good practices to preservice teachers. Their results revealed the need for distinction between mobile learning and mobile usage, as well as the necessity for exploring the pedagogical potentials of mobile technologies (Schuck et al., 2013). In another study, tablets were integrated into literacy coaching practices and classroom observations (Bates & Martin, 2013). Hargis et al. (2013) conducted observations, interviews, and surveys with faculty members who received training on iPads and found that iPads supported student-centered teaching. In another study, teacher educators collaboratively redesigned their coursework to emphasize “(a) assessing the value of mobiles in instruction, (b) pedagogical approaches, (c) content creation, (d) evaluation of apps, and (e) learner impact” (Herro et al., 2013, p. 35).

While faculty development is considered critical in adopting mobile learning practices in teacher education, only a limited number of studies examined teacher educators’ practice and training about mobile learning. Teacher education programs can initiate mobile learning systems in collaboration with K-12 schools by implementing “visionary leadership, professional development, scalable pilot programs, and adequate resources” (Herro et al., 2013, p. 36).

Overall, this review captured the methods for integrating mobile learning into teacher education contexts. Mobile in teacher education contexts aimed to (a) teach teachers how to integrate mobile tools into their classrooms, and to (b) enhance teacher learning with mobile learning. Figure 3 summarizes these methods.

**Figure 3. Methods for integrating mobile learning into teacher education**

- **Teacher training about mobile learning**
  - Hands-on explorations of mobile technologies
  - Developing mobile lesson plans
  - Micro-teaching mobile lessons
  - Enacting mobile lessons in the classrooms
  - Reflecting on mobile lessons
  - Planning mobilized curriculum
  - Attending to communities of practice

- **Teacher training with mobile learning**
  - Collaborating with peers, colleagues, teacher educators and supervisors
  - Accessing teacher education content anytime anywhere
  - Reflecting on teaching
  - Sharing classroom practice
  - Using peer feedback
  - Customizing teacher training content
  - Observing teachers real time
  - Assessing performance
Conclusions

This systematic review of 37 articles on mobile learning and teacher education is timely in light of growing interest in mobile learning and a lack of syntheses in the context of teacher education. Findings are drawn as well as the approaches and strategies for implementing mobile learning and mobile tools in different teacher education contexts. First, the review revealed that the number of articles published has significantly increased over the last five years, with contributions from researchers around the world. This trend is consistent with other review findings on mobile learning (e.g., Hwang and Tsai, 2011; Wu et al., 2012). Second, there was a scant report of theoretical and conceptual perspectives. Third, survey studies revealed that perceptions, attitudes and usage patterns among teachers varied. Fourth, mobile learning was reported as mainly beneficial in teacher education contexts. Fifth, while there are notable exceptions, challenges or issues to integrate a mobile learning component in teacher education programs and curriculum were duly noted. Finally, several pedagogical affordances support mobile learning integration into teacher education settings. As educators begin to understand the potential of mobile learning in education, the role of teachers and teacher educators in integrating mobile devices becomes essential in addressing students’ learning needs across several disciplines. This study presents findings and recommendations to help researchers, teacher educators, and policy makers develop research-informed guidelines and theoretical models to suggest methods on how to integrate mobile learning into teacher education.

Recommendations for future research and practice

This systematic review revealed a number of critical recommendations for those who plan to investigate the advantages of mobile learning for preservice and in-service teachers and integrating these technologies into teacher education contexts.

Transforming teacher education practices with theoretically sound approaches

Teacher educators need to go beyond the tools’ potential to explore pedagogical benefits of mobile learning within their own content areas. This change will help preservice and in-service teachers realize the pedagogical advantages of mobile learning that may shift their perspectives toward the integration of mobile devices into their teaching environments. While a number of studies presented the benefits of using mobile devices within field experiences, more research is needed to understand its unique applications as well as its impact on mentor teachers’, preservice teachers’, and teacher educators’ successful implementation of school-based practicums. To understand the sustained effect of mobile learning initiatives in teacher education programs, teacher educators and researchers also need to conduct longitudinal studies observing classrooms over time as well as enacting pedagogical approaches to mobile devices within actual classroom settings (Price et al., 2014).

Investigating additional strategies for mobile learning integration and expanding data corpus on mobile learning

Further research needs to investigate how different types of mobile learning projects, time requirements, and site contexts have an impact in the way teachers develop outcomes (e.g., content knowledge, pedagogical skills) and how different types of reflection, observation, design, planning, engagement, and assessment activities effect these outcomes. Based on this synthesis, it is also recommended that interested researchers expand the data corpus on mobile learning and establish its relationship to desired teacher outcomes, such as technological pedagogical content knowledge (TPACK) within mobile learning contexts and address issues endemic to teacher development and education in the teacher preparation phase.

Using varied research methodologies with diverse and larger samples and reports on trustworthiness

The majority of the research conducted on mobile learning and teacher education presents best practices and case studies conducted within contexts with limited scope and small sample size. While these best practices reveal critical findings, future empirical research must follow other methodological routes such as design-based research to develop
theories within practice and ethnography to understand how mobile learning interacts with social and cultural dynamics in teacher education contexts. The field also calls for studies with diverse and larger samples with rich descriptions of context and trustworthiness to make the findings transferable.

Developing new models for teachers’ professional development using mobile learning

Research on mobile learning and teacher education technology has mostly investigated teachers learning about mobile technologies rather than learning with them. While a limited number of studies have looked at learning with mobile devices (Aubusson et al., 2009), more research is needed to understand how teachers and teacher educators’ professional development can be supported with mobile learning. Additional research may also consider how the immersion of teacher educators into CoPs or PLCs helps them engage in a professional conversation about the integration of mobile learning within teacher education contexts, such as sharing stories and best practices, engaging in collaborative resource creation, and helping each other through a mentor or peer support system (Herro et al., 2013).

A systemic investigation of mobile learning in preservice teacher education

Greater insight into research on mobile learning in specific teacher education contexts has potential to support more system-wide adoptions of mobile learning, where more research is needed. While the research on preservice teacher education and mobile learning mainly included the investigation of mobile learning as an approach within individual teacher education courses, further research continues to be needed on the systemic investigations of mobile learning within entire teacher education programs. Different ways for mobile learning integration into teacher education can be examined, such as the infusion in practica, special courses (e.g., methods or classroom management), and throughout the teacher education program. Understanding the potential impact of mobile learning that is integrated into the entire teacher education programs and associated challenges and benefits are critical to gaining greater insight into the purposes of various phases of teacher education and the role of mobile learning in each.

References


*References marked with an asterisk indicate the articles used in the review.
Methodological Issues in Mobile Computer-Supported Collaborative Learning (mCSCL): What Methods, What to Measure and When to Measure?

Yanjie Song
Department of Mathematics and Information Technology, The Hong Kong Institute of Education // ysong@ied.edu.hk

ABSTRACT
This study aims to investigate (1) methods utilized in mobile computer-supported collaborative learning (mCSCL) research which focuses on studying, learning and collaboration mediated by mobile devices; (2) whether these methods have examined mCSCL effectively; (3) when the methods are administered; and (4) what methodological issues exist in mCSCL studies. It attempts to bring to light methods more conducive to examining the effectiveness of mCSCL and thus to sustain the practices. The research findings reveal a variety of methodological issues that need to be addressed. Comparison is made to the findings in CSCL research and other studies leveraged by mobile technologies. Potential ways to investigate the effectiveness of mCSCL practices are proposed.

Keywords
mCSCL, Methods, Measures, Methodological approach

Introduction

Computer-supported collaborative learning (CSCL) research is concerned with studying how people can learn together with the help of computers as an emerging field of learning sciences (Stahl, Koschmann, & Suthers, 2006); while Mobile Computer-Supported Collaborative Learning (mCSCL) research focuses on learning and collaboration mediated by mobile devices (e.g., Zurita & Nussbaum, 2004). Stahl et al. (2006) maintain that it is a challenging task to combine the two elements “computer support” and “collaborative learning” to effectively enhance learning that CSCL is designed to address. Yet, it is an even more challenging task to understand how to combine the ideas of “mobile computer support” and “collaborative learning” to advance learning and collaboration in different settings and modes because mCSCL involves many changing practices partly due to the unique technology characteristic of “mobility” and the dynamically re-constructed context for interaction and learning (Looi, Wong, & Song, 2012; Arrigo, Kukulska-Hulme, Arnedillo-Sánchez, & Kismihok, 2013). Various tools and systems have been increasingly developed on, or integrated into, mobile devices for mCSCL. These activities have been carried out across different spaces physically, socially and virtually. However, despite a number of studies reporting the benefits gained from the implementation of mobile tools or systems for mCSCL, little impact has been observed on actual educational practices outside the context of research investigations (Roschelle et al., 2010). To uncover the “black box” of mCSCL research, it is crucial to explore current mCSCL research designs, especially current methodology approaches to mCSCL practices in order to understand how groups and individuals make sense of situations and construct knowledge supported by mobile technologies, and what methodological issues are unique to mCSCL practices. Thus, this study attempts to conduct a systematic review of methodological approaches found in the mCSCL research literature in order to bring to light potential methods more conducive to examining the effectiveness of mCSCL practices, and to uncover methodological design issues in mCSCL practices to be addressed in future research. This, in turn, will help sustain the practices.

This paper first provides a description of how the literature review was carried out. Second, the results of the review are presented followed by discussions of methodological issues and potential directions for future mCSCL research. Finally, conclusions are drawn.

Method

To understand mCSCL practices, a systematic review was carried out. A systematic review refers to a review of the literature based on explicit, rigorous and transparent methodology (Coffield, Moseley, Hall, & Ecclestoneaimeis, 2004). This study systematically reviewed and synthesized the relevant literature on mCSCL research to unpack the methodological approaches adopted in the studies (Coffield et al., 2004; Wong & Looi, 2011; Wu, Wu, Chen, Kao, Lin, & Huang, 2012). In general, the first phase of a systematic review is framing the research questions regarding
which the review is carried out. This is followed by a thorough search of the relevant literature, followed by a check of the criteria in selecting articles that meet the review purposes. The second phase is the review process. Finally, the third phase is to write the report of the review. The review is intended for researchers who are interested empirical studies in the area of mobile/handheld educational applications, especially methodological issues in mCSCL practices.

The research questions are:
- What methods are utilized in mCSCL research?
- When are the methods administered?
- Have these methods examined mCSCL effectively?
- What issues do the methodological approaches have in existing mCSCL studies?

Selection criteria

To address the research questions, a set of criteria drawn from the literature (e.g., Gress et al., 2010; Hsu & Ching, 2013; Wong & Looi, 2011) was used in selecting articles that met the review purposes. These criteria are: (a) mCSCL studies that have addressed group collaboration supported by mobile technologies; (b) empirical studies, including case studies and evaluation studies with empirical evidence; (c) studies that have explicit research questions/statements; (d) studies that include handheld devices which refer to mobile devices that are small, handheld computing devices, typically having a display screen with touch input (http://en.wikipedia.org/wiki/Mobile_device), and are considered having strong a “mobility” nature such as PDAs, Smartphones, iPad, and mobile phones; and (e) studies published in refereed journals. In this review, studies that have used laptop computers to support collaboration are excluded because they do not belong to “handheld devices.” In addition, studies that are concerned mainly with conceptual frameworks, literature reviews, and technical infrastructures are beyond the focus of this review, and are excluded.

Identification of eligible mCSCL studies

The literature search and review underwent three stages. First, an extensive literature search was conducted in 33 major refereed academic journals related to technology-enhanced teaching and learning research, starting from the year 2000 when mobile educational research was at its infancy to February 2014 when this review work initiated. As this review focuses on studies of mCSCL that provide empirical evidence, refereed journals emphasizing mobile technological architecture design are excluded. In addition, articles from conferences and workshops are not included due to issues of inaccessibility or similarity of conference reports and journal articles. The search used the key words “mCSCL,” or “mobile computer-supported collaborative learning,” and “mobile” and “collaborative learning.” These journals were: Australasian Journal of Educational Technology, British Journal of Educational Technology, CALL (Computer-Assisted Language Learning), Computers & Education, Computers in Human Behavior, Educational Technology & Society, Educational Research Review, Educational Review, Educational Media International, Educational Technology Research and Development, ELT Journal, Instructional Science, International Journal of Computer-Supported Collaborative Learning, International Journal on E-Learning, Interactive Learning Environments, International Journal of Lifelong Education, International Journal of Science Education, International Journal of Technology in Teaching and Learning, Internet and Higher Education, Journal of Computers in Mathematics and Science Teaching, Journal of Computer Assisted Learning, Journal of Educational Computing Research, Language Learning & Technology, Learning and Instruction, RECALL, Research and Practice in Technology Enhanced Learning, Review of Educational Research, Technology, Pedagogy and Education, The Asia-Pacific Education Researcher, The Interdisciplinary Journal of E-Learning and Learning Objects, TechTrends, and The Journal of Learning Sciences. As of February 2014, the search of the publications in these journals yielded 122 results, among which, 31 articles met the criteria. A second round of search was conducted in GoogleScholar using the same key words as the literature search in refereed journals. The first ten pages of search results on GoogleScholar of each key word combination were viewed, from which 3 more articles were added to the pool. Finally, a third round of search used the snowball sampling approach (Gao, Luo, & Zhang, 2012) by scanning references cited in previous selected articles. One more article was identified and added to the pool. As a result, 35 papers were identified as eligible articles for the review and analysis (see Figure 1).
The distribution of eligible articles is spread among 15 journals which are shown in Table 1.

Table 1. Distribution of eligible articles in all referred journals

<table>
<thead>
<tr>
<th>Journal name</th>
<th>No.</th>
<th>Journal name</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computers &amp; Education</td>
<td>9</td>
<td>Computers in Human Behavior</td>
<td>1</td>
</tr>
<tr>
<td>Educational Technology &amp; Society</td>
<td>7</td>
<td>Language Learning &amp; Technology</td>
<td>1</td>
</tr>
<tr>
<td>Interactive Learning Environments</td>
<td>4</td>
<td>Computer Assisted Language Learning</td>
<td>1</td>
</tr>
<tr>
<td>British Journal of Educational Technology</td>
<td>3</td>
<td>Journal of Science Education and Technology</td>
<td>1</td>
</tr>
<tr>
<td>Journal of Computer Assisted Learning</td>
<td>3</td>
<td>Asia-Pacific Education Researcher</td>
<td>1</td>
</tr>
<tr>
<td>Journal of the Learning Sciences</td>
<td>1</td>
<td>Educational Media International</td>
<td>1</td>
</tr>
<tr>
<td>International Journal of Computer-Supported</td>
<td>1</td>
<td>Research and Practice in Technology Enhanced Learning</td>
<td>1</td>
</tr>
<tr>
<td>Collaborative Learning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of mCSCL studies

This analysis has been illuminated by the review studies and commentary on examining measurement and assessment in CSCL and mCSCL review studies (e.g., Hmelo-Silver & Bromme, 2007; Gress, Fior, Hadwin, & Winne, 2010; Looi et al., 2012). According to Hmelo-Silver and Bromme (2007), measurement and methodological approaches are dependent on the theoretical framework being used and the research questions being asked. Measurement in CSCL can take one of the three forms: assessing the individual about the individual’s experience (assessing the individual learning process and outcomes), assessing the individual about the group’s experience (assessing the individual learning process and outcomes contributed to the group collaborative learning), and assessing the group as a whole (assessing group’s learning process and outcomes) (Gress et al., 2010; Hmelo-Silver & Bromme, 2007). Measurement in CSCL includes observing, capturing and summarizing both individual and group behaviors, from which researchers infer learning processes and outcomes. Factors affecting measurement in CSCL consist of individual differences, context, tool use, collaborative activities, and researchers’ different theoretical backgrounds (Gress et al., 2010). Because of the mobility nature of mobile devices, collaborative learning may occur in constantly changing contexts (e.g., from physical to virtual, from individual to social, and from informal to formal learning spaces). Thus, it is a very demanding task to capture the students’ learning process. For example, Song (2014) reported on a study of a collaborative science inquiry on the topic “Anatomy of fish” supported by mobile devices in a primary school. The pedagogical design included learning activities across different spaces – at home where students were asked to explore information about fish in a web market and online; in the school lab where students observed real fish brought by the teacher for each group; and in class and on a social network platform where students shared their findings and reflections. Hence, students learning experiences were not constrained to one place but occurred in constantly changing situations, which made tracking students’ learning trails a challenging task. To understand mCSCL practices, we need to examine holistically and “re-construct” (Avouris et al., 2007, p. 248) learning scenarios occurring in different contexts. To demonstrate how the mCSCL environment and tools benefit learning, it is essential to focus on methodological approaches (instruments/techniques), measures and analysis (Gress et al., 2010; Hmelo-Silver & Bromme, 2007). Based on the attributes of CSCL research, Gress et al. (2010) identify a coding scheme consisting of three key methodological aspects to be addressed: measures (e.g., efficacy and knowledge construction), methods (e.g., questionnaires, interviews, observations and discourse analysis), and measurement timing (e.g., assessment/evaluation timing before, during or after the collaborative task) in the CSCL context. Premised on Gress et al. ’s (2010) coding scheme, to answer the present study’s research questions, an analysis framework was developed for the purposes of reviewing and analyzing mCSCL research. In the framework,
a new dimension “context of studies” was added by taking the “mobility” nature of mCSCL into account. Also, another element, “research design and foci,” was added. The framework consists of five elements: context of studies, research design and foci, methods, measures, and timing of measurement.

(a) Context of studies: This includes participants, sample size, duration of intervention, domain areas, adopted mobile technology/tools, and settings [in-class, planned/emerging; out-of-class, planned/emerging learning environments (Chen, Seow, So, Toh, & Looi, 2010; So, Kim, & Looi, 2008), and in-and-out-of-class mixed learning environments (Song, 2014)] which will be elaborated in a later section.

(b) Research design and foci: This relates to the research methodology adopted and research aims or statements in the studies. For example, Laru, Järvelä, and Clariana (2012) state that “The aim of the analysis was to identify and compare top- and low-performing dyads/triads in order to reveal the differences regarding their co-construction of arguments while creating knowledge claims” (p. 1).

(c) Methods: this refers to all the instruments such as questionnaires, surveys, discourse analysis, content analysis, and artifact analysis used in mCSCL research and techniques reported as used in the articles such as interviews, observations, discussions, and process data.

(d) Measures: these refer to dependent variables/subjects that are measured in the studies with the above instruments. Based on the research focus, the measures were more precisely identified. For example, the measure adopted in Laru et al. (2012) was co-construction of arguments.

(e) Timing of the measurement (before, during or/and after): this refers to the measuring/assessment timing (before, during, or after mCSCL practice) in the studies (Gress et al., 2010) (e.g., assessing students’ performance after collaboration in Laru et al., 2012). In addition, forms of measurement (assessing the individual about the individual’s experience, assessing the individual about the group’s experience, and/or assessing the group as a whole) are also examined (e.g., assessing the group as a whole in Laru et al., 2012).

Guided by the analysis framework, a content analysis of the 35 articles was conducted. The process of analysis consisted of four steps. The analysis framework together with the four-step analysis process across the entire study to address the four research questions is shown in Figure 2.

Figure 2. Analysis framework and analysis process of the 35 eligible mCSCL studies

In the first step, all the individual articles were coded based on the 5 elements of the analysis framework. A preliminary sorting was worked out. Then a further categorization of the articles was made to verify the characteristics of each element in the framework: (a) the context of studies were categorized into participants (primary, secondary, tertiary education or others), sample sizes (< 10, 10–50, 51–100, and > 100), duration of
intervention (1–5 days, 1–4 weeks, 5–8 weeks and > 8 weeks), domain areas (e.g., maths, science, etc.), adopted mobile technology/tools (e.g., smartphones, mobile learning systems/apps, others), and settings (in class, planned; out of class, planned, in and out of class, mixed, and others); (b) research design and focus (e.g., methods and the aims/objectives of studies); (c) methods (instrument/techniques such as questionnaires, surveys, interviews, observations, discussions and dialogue, feedback); (d) measures (dependent variables/subjects that are measured in the studies with the instruments, such as learning performance and collaborative learning behaviors/patterns); and finally (e) timing of the measurement (before, during, or after mCSCL practices).

The second step focused on identifying common themes of mCSCL practices and effects in the five dimensions across 35 articles to address the first two research questions (Q1: what methods? And Q2: when being measured?). Two researchers, based on the categories coded in the first step, independently coded the themes in the studies of the articles, and then compared and discussed the themes to reach consensus. The third step, based on the first two steps’ work, addressed the third research question (Q3: Have these methods examined effectively?). Finally the fourth step, based on the previous three steps’ work, addressed the fourth research question (Q4: What issues?). An example of the coding process is shown in Table 2.

Table 2. Example of the coding process in reaching consensus on the study by Lan, Sung, & Chang (2009)

<table>
<thead>
<tr>
<th>Coder 1</th>
<th>Research focus (RF)</th>
<th>Measures (M)</th>
<th>Methods (I)</th>
<th>Timing (T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve collaborative learning approaches (#RF2)</td>
<td>Collaborative behaviors (#M2); metacognitive strategies (#M3); prior knowledge (#M5)</td>
<td>Observations (#I2); performance and products (#I6)</td>
<td>Before (TB), during (TD), after (TA)</td>
<td></td>
</tr>
<tr>
<td>Evaluate the effectiveness of learning system (#RF3)</td>
<td>Learning performance (#M1); metacognitive strategies (#M3)</td>
<td>Observations (#I2); process data (#I4); performance and products (#I6);</td>
<td>Before (TB), during (TD), after (TA)</td>
<td></td>
</tr>
</tbody>
</table>

Consensus (after discussion)
Evaluate the effectiveness of learning system (#RF2)
Learning performance (#M1); Collaborative behaviors/patterns (#M2); Metacognitive strategies (#M3)
Observations (#I2); Process data (#I4); performance and products (#I6); | Before (TB), during (TD), after (TA) |

Results

Research question (1): What methods are utilized in mCSCL research?

To answer this question, Context of studies, and research design and foci were first identified, from which methods of mCSCL learning practices were distilled and categorized into 7 types. Further methods utilized in mCSCL practices were discerned in the course of content analysis of the 35 studies.

Context of studies

The context of the 35 studies are presented to provide a general picture of mCSCL practices. The context of a study influences the measurement of collaborative practices (Hmelo-Silver & Bromme, 2007). This is especially true in mCSCL practices (e.g., Hwang et al., 2011; Laru et al., 2012; Looi et al., 2012). The contexts of the 35 studies were categorized in terms of (a) participants, (b) sample size, (c) duration of intervention, (d) domain areas, (e) adopted mobile technology/tools, and (f) settings.

(a) Participants: Out of the 35 studies, 37 % was concerned with participants from primary schools, and 34 % and 29% were concerned with participants from tertiary education and secondary schools respectively.

(b) Sample size: Out of the 35 studies, 21 (60%) had sample sizes ranging from 10-50, and 2 (5%) had sample sizes smaller than 10, and 1 (3%) study did not mention its sample size.

(c) Duration of intervention: Almost half of the 35 studies (48%) had a research intervention ranging from 1 to 4 weeks. Ten studies (28%) had 1 to 5 days’ research interventions, which accounts for the largest percentage among all studies; and 2 studies (6%) did not report how long the intervention was involved.
(d) Domain areas: The domain area of 12 (29%) studies was science, followed by language (19%), IT related studies (19%), maths (17%), and others (7%). It is noted that all articles regarding IT related domain areas were conducted in tertiary education. In addition, 5 studies involved two to three subject areas.

(e) Mobile technology/tools: 13 studies (36%) chose smartphones as the mobile devices to support students’ collaborative study, followed by PDAs (34%), mobile phones (14%), tablets including iPad (8%), and others (8%).

(f) Settings: Researchers of mobile educational applications such as Chen et al., (2010) and So et al., (2008) classified the spaces where mobile learning happens into four types: (1) planned learning in class (e.g., Chinese language collaborative learning included in the curriculum in Wong et al., 2011); (2) planned learning out of class (e.g., field trips for science learning included in the curriculum in So et al., 2009); (3) emergent learning in class (e.g., new vocabulary look-up using mobile device in class that is not included in the planned activities in Song & Fox, 2008); and (4) emergent learning out of class (e.g., taking pictures using mobile devices out of self interest/motivation in Wong, Chen, & Jan, 2012). In this review, it is found that the majority of studies (65%) carried out the research in “planned learning in class,” followed by “planned learning out of class” (20%) and “planned learning in both in-and out-of-class” (15%). No research has reported findings regarding unplanned, “emergent learning out of class.”

Research design and foci

The research foci of the 35 studies were classified into five categories: (a) to improve the collaborative activity/process/learning using CSCL systems (12 studies, 34%); (b) to evaluate the effectiveness of learning systems/tools (10 studies, 29%); (c) to improve collaborative learning strategies/approaches (8 studies, 23%); (d) to explore the educational potential of learning systems/tools (4 studies, 11%); and (e) to discover collaborative learning patterns (1 study, 3%).

It is noted that among the 35 studies, 15 of them (44%) adopted experimental or quasi-experimental design, followed by 6 (17%) studies employing a case study approach, 5 (14%) studies doing learning system evaluation, and five studies (11%) using design-based research approach.

Measures

Ten types of measures were identified (see Table 3). They are: learning performance (19%); collaborative behaviors/patterns (18%); prior knowledge/skills (16%); student satisfaction/attitude/ perception towards learning system/tool (16%); metacognitive strategies (11%); process of collaborative investigations (10%); perception of learning skills (problem solving/ inquiry skills, collaborative skills) (3%); participation in collaborative activities (3%); self-efficacy and the local culture identity (2%); and affordances and limitations of collaborative learning system/tool (2%). Some of the studies had more than one measure to achieve their research aims.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Studies</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative behaviors/ patterns</td>
<td>Cortez et al., 2005; Zurita &amp; Nussbaum, 2007; Rogers &amp; Price, 2008;</td>
<td>12</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>Roschelle et al, 2010; Chang &amp; Hsu, 2011; Hwang et al., 2011; Laru et al., 2012; Ryu &amp; Parsons, 2012; Lin et al., 2013; Sung et al., 2013; Song, 2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning performance</td>
<td>Zurita &amp; Nussbaum, 2004; Lan et al., 2007; Zurita &amp; Nussbaum, 2007;</td>
<td>11</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>Lan et al., 2009; Liu et al, 2009; Capponi et al, 2010; Boticki et al, 2011; Wong et al., 2011; Lan et al., 2012; Timmis, 2012; Wong et al., 2012, Lin et al., 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior knowledge/skills</td>
<td>White, 2006; Zurita &amp; Nussbaum, 2007; Lan et al., 2009; Roschelle et al, 2010; Chang &amp; Hsu, 2011; Hwang et al., 2011; Laru et al., 2012; Lin et al., 2013; Sung et al., 2013; Song, 2014</td>
<td>10</td>
<td>16%</td>
</tr>
<tr>
<td>Student satisfaction/attitude/ perception</td>
<td>Wei &amp; Chen, 2006; Zurita &amp; Nussbaum, 2007; Huang et al., 2008; Huang et al., 2009; El-Bishouty et al., 2010; Chang &amp; Hsu, 2011; Echeverria et al., 2011; Hwang et al., 2011; Wong et al., 2011; Lan et al., 2012</td>
<td>10</td>
<td>16%</td>
</tr>
</tbody>
</table>
Methods adopted in mCSCL research

The methods utilized in mCSCL research were coded (Gress et al., 2010) and classified into 7 types for the measurement of the effectiveness of mCSCL. They are summarized in Table 4.

Table 4. Methods, instruments/techniques and distribution among the 35 studies

<table>
<thead>
<tr>
<th>Methods</th>
<th>Instruments/techniques</th>
<th>No. &amp; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report</td>
<td>Questionnaires, surveys, summative project reports</td>
<td>16 (18%)</td>
</tr>
<tr>
<td>Interviews</td>
<td>Discussions between researchers, teachers and students</td>
<td>11 (12%)</td>
</tr>
<tr>
<td>Observations</td>
<td>All methods of visually examining and documenting actions and utterances of participants, either directly or by videotape recording</td>
<td>13 (15%)</td>
</tr>
<tr>
<td>Process data</td>
<td>Estimates of time, frequency, and sequence, as well as tracing data which examined participants’ actions via the computer during the collaborative tasks</td>
<td>14 (16%)</td>
</tr>
<tr>
<td>Discussions &amp; dialogues</td>
<td>Engaged purposeful conversation and/or verbal expressions coded as either asynchronous or synchronous communication.</td>
<td>6 (7%)</td>
</tr>
<tr>
<td>Performance &amp; products</td>
<td>All output produced by participants’ collaborative activities</td>
<td>24 (27%)</td>
</tr>
<tr>
<td>Feedback</td>
<td>Feedback from participants, teachers, researchers</td>
<td>4 (5%)</td>
</tr>
</tbody>
</table>

Some of the studies employed more than one type of method (instruments/techniques) ranging to up to six types. Percentages are shown in Figure 3.

Figure 3. Number of types of methods adopted in the 35 studies
Research Question (2): When are the methods conducted?

The timing of measurement was classified into 3 types in the 35 studies: before, during and after the mCSCL practices. The results show that 10 studies (18%) did the assessment before the mCSCL practices, and 23 studies (42%) did the assessment during the mCSCL practices, and 22 studies (40%) did the assessment after the mCSCL practices.

Further, an investigation was carried out to discern the patterns of assessment at different timings with varied methods as shown in Table 5. Table 5 displays that the assessment before the mCSCL practices emphasises (a) prior knowledge/skills; (b) student satisfaction/attitude/ perception towards learning system/tool, and (c) baseline information about self-efficacy and the local culture identity.

The assessment during the mCSCL practices focused on (a) collaborative behaviors/ patterns, (b) metacognitive strategies; (c) process of collaborative investigations; (d) participation in collaborative activities; and (e) affordances and limitations of collaborative learning system/tool.

Finally, the assessment after the mCSCL practices centered on (a) learning performance; (b) student satisfaction/attitude/ perception towards learning system/tool; (c) perception of learning skills (problem solving/ inquiry skills, collaborative skills; (d) metacognitive strategies, and (e) self-efficacy and the local culture identity.

<table>
<thead>
<tr>
<th>Timing</th>
<th>Measures</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>prior knowledge/skills</td>
<td>White, 2006; Zurita &amp; Nussbaum, 2007; Lan et al., 2009; Chang &amp; Hsu, 2011; Laru et al., 2012; Sung et al., 2013; Song, 2014</td>
</tr>
<tr>
<td></td>
<td>Student satisfaction/attitude/ perception towards Learning system/tool</td>
<td>Chang &amp; Hsu, 2011; Hwang et al., 2011; Lin et al., 2013</td>
</tr>
<tr>
<td></td>
<td>baseline information about self-efficacy and the local culture identity</td>
<td>Sung et al., 2013</td>
</tr>
<tr>
<td>During</td>
<td>collaborative behaviors/ patterns</td>
<td>Zurita &amp; Nussbaum, 2004; Lan et al., 2007; Lan et al., 2009; Capponi et al., 2010; Wong et al., 2011; Timmis, 2012; Wong et al., 2012; Lin et al., 2013</td>
</tr>
<tr>
<td></td>
<td>metacognitive strategies</td>
<td>Lim &amp; Wang, 2005; White, 2006; Cortez et al., 2009; Nussbaum et al. 2009; Boticki et al., 2011; Lan et al., 2012</td>
</tr>
<tr>
<td></td>
<td>process of collaborative investigations</td>
<td>Colella, 2000; Rogers &amp; Price, 2008; So et al., 2009; Ryu &amp; Parsons, 2012; Song, 2014</td>
</tr>
<tr>
<td></td>
<td>participation in collaborative activities</td>
<td>Wei &amp; Chen, 2006; Liu et al, 2009; El-Bishouty et al., 2010</td>
</tr>
<tr>
<td></td>
<td>affordances and limitations of collaborative learning system/tool</td>
<td>Dunleavy et al., 2009</td>
</tr>
<tr>
<td>After</td>
<td>Learning performance</td>
<td>Cortez et al., 2005; Zurita &amp; Nussbaum, 2007; Rogers &amp; Price, 2008; Roschelle et al, 2010; Chang &amp; Hsu, 2011; Hwang et al., 2011; Laru et al., 2012; Ryu &amp; Parsons, 2012; Lin et al., 2013; Sung et al., 2013; Song, 2014</td>
</tr>
<tr>
<td></td>
<td>Student satisfaction/attitude/ perception towards Learning system/tool</td>
<td>El-Bishouty et al., 2010; Huang et al., 2008; Huang et al., 2009; Echeverría et al., 2011; Lan et al., 2012; Wong et al., 2011</td>
</tr>
<tr>
<td></td>
<td>Perception of learning skills (problem solving/ inquiry skills, collaborative skills</td>
<td>Wei &amp; Chen, 2006; Sánchez &amp; Olivares, 2011; Song, 2014</td>
</tr>
<tr>
<td></td>
<td>Metacognitive strategies</td>
<td>White, 2006; Lan et al., 2009; Liu et al., 2009</td>
</tr>
<tr>
<td></td>
<td>Self-efficacy and the local culture identity</td>
<td>Sung et al., 2013</td>
</tr>
</tbody>
</table>
Some of the studies did the assessment across two or three timings as shown in the above table. 12 studies (36%) assessed students’ learning process and skills during mCSCL practices, 8 studies (22%) assessed students’ learning after the mCSCL practices; while 5 studies (14%) assessed students’ learning in all ranges of timings (before, during and after), 5 studies (14%) did the assessment during and after mCSCL practices, and 5 studies (14%) did the assessment before and after the practices.

How was the assessment conducted among the participants of the 35 studies for mCSCL practices? Two forms of measurement were identified: individual about group and group as a whole. Some of the studies employed both the forms to assess mCSCL practices. About 16 studies (46%) assessed collaboration via group as a whole; 10 studies (28%) assessed collaboration via individual about group and group as a whole studies, and 9 (26%) assessed collaboration via individual about group. None assessed individual about individual.

Research question (3): Have these methods examined mCSCL effectively?

“Premised on the “Analysis framework and analysis process of mCSCL studies” (see Figure 2) developed in this study research question (3) was addressed by (a) grouping the methods adopted in these studies with the timing of the measurement (Gress et al., 2010) resulting from research questions (1) and (2), and (b) contextualizing the 35 eligible studies by referring to the results obtained from the context of studies.

Grouping the methods with timing of measurement

First, the 7 types of methods, which were adopted in the 35 studies 88 times, were grouped together with the timing of the measurement. The frequency of the methods is shown in Table 6.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Measurement timing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>During</td>
</tr>
<tr>
<td>Self-report</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Interviews</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Observations</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Process data</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Discussions and dialogues</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Performance and products</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Feedback</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total n</td>
<td>12</td>
<td>36</td>
</tr>
</tbody>
</table>

*% n = Percentage of methods in each type of timings (before, during, and after)
**% N = Percentage of methods among all types of timings (before, during, and after)

It is noted from Table 6 that only 12 methods (13% of the total methods) were concerned with assessment before the mCSCL practices. The measures mainly concentrated on using performance and products (10%) to measure prior domain knowledge and collaborative skills; while 3% of the total methods used self-report (e.g., questionnaires) to measure student satisfaction/attitude/perception towards learning system/tool identity (e.g., Chang & Hsu, 2011; Hwang et al., 2011), and baseline information about self-efficacy and the local culture (Sung et al., 2013).

During the mCSCL practices, it was found that 36 methods (42% of the total methods) were used, among which, 13 methods (15%) used observation, 14 methods (16%) used process data followed by 6 methods (7%) of discussions and dialogues and 3 methods (4%) of feedback. These methods were used to assess collaborative behaviors/patterns, metacognitive strategies, process of collaborative investigations, participation in collaborative activities and affordances and limitations of collaborative learning system/tool.

After the mCSCL practices, it was found that 40 methods (45% of the total methods) were used, among which, 13 methods (15%) adopted self-report and 11 methods (12%) used interviews, 15 methods (17%) used performance and
products and 1 method (1%) used feedback to assess: learning performance, student satisfaction/attitude/ perception towards learning system/tool, perception of learning skills such as problem solving/ inquiry skills, collaborative skills, metacognitive strategies, and self-efficacy and the local culture identity.

**Contextualizing the 35 eligible studies**

This section presents the results from contextualizing and comparing 35 eligible studies in order discern methodological patterns.

It was observed that a fairly low number of methods (13%) were conducted before the mCSCL practices and a comparatively high number of methods (45%) were administered after the mCSCL practices although 44% of the 35 studies adopted experimental designs. In addition, it was found that the majority of the 35 studies (21 studies, 60%) had sample sizes ranging from 10-50, and 2 studies (5%) only had sample sizes smaller than 10. Thus, the significance of the results from assessment before and after mCSCL practices is challenged.

Next, it is noted that although 42% of the total methods were employed during mCSCL practices, the intervention duration tended to be too short, ranging from 1 day to 4 weeks among 48% of the total 35 studies although all the concerned studies reported positive results. This poses a challenge against the sustainability of the collaborative process in these studies in the long run. In addition, many of the studies lacked a clear coding framework and chose only short episodes of the process data. Thus, the measures are ambiguous.

Moreover, based on the publication years, the context of studies and methodological approaches were compared across the 35 studies. Out of the 17 studies published during the years 2000-2009, 71% adopted PDAs/Palms or mobile devices with wireless connection (including intranet connection), and the rest used Tablet PCs, smartphones or other wearable devices with wireless connection; while out of the 18 studies published during 2010-2014, 56% adopted smartphones or iPads, and the rest used PDAs or mobile phones. With the evolution of mobile devices, the methodological approaches in the studies also evolved: (a) before 2000-2009, the context of the studies was either in-class or out-of-class planned activities; from 2010, the context of some studies included both in-class and out-of-class planned activities (e.g., Hwang et al., 2011; Song, 2014); ((b) from 2010, more and more studies focused on examining contextualized just-in-time content creation and meaning making in learners’ collaborative learning experiences in real life encounters (e.g., Ryu & Parsons, 2012; So et al., 2012; Song, 2014; Wong et al., 2012), and (c) from 2010, design-based research (Collins, Joseph, & Bielaczyc, 2004) was on the rise (e.g., Roschelle et al., 2010; Wong et al., 2012), which attempts to combine theory-driven design with empirical analyses of practices in real settings. It is noted that design-based research is not unique to mCSCL but was developed and practiced by a group of learning scientists in early 21st century. Despite these changes in mCSCL practices, it is also to be noted that existing methods (e.g., self-report, interview, observation) and measures (learning performance, collaborative behaviors/patterns) were widely employed across these studies.

Research question (4): What issues do the methodological approaches have in existing mCSCL studies?

This research question is addressed based on the results from the first 3 research questions as well as the analysis framework (see Figure 2). Issues of the methodological approaches are pinpointed and summarized in the following seven aspects.

**Lack of methods before mCSCL practices**

The results of the review study revealed that only a small number of studies conducted measurement before the mCSCL activity, and these methods before mCSCL practices focus on using self-report to measure prior knowledge or skills, and students’ perception/attitude towards the collaborative learning system or tool. Other methods such as interviews and focus group discussions were not tapped into for baseline information in the 35 studies. The form of assessment was restricted to individual about group.
Lack of methods of examining mCSCL processes

Although less than half of the 35 studies employed measurement during the mCSCL activity and assessed collaboration via the group as whole, the types of methods were limited largely to interview, observation, process data and discussion and dialogues. The descriptive approach, if used properly for fine-grained analysis, can provide rich pictures of interactions (Stahl, 2006). However, in many cases, only short episodes of collaborative discussion data were selected for the analysis without clear coding schemes, the process of collaboration was not clearly revealed, and sometimes the methods adopted were vague. This may be due to difficulties in capturing mobile learners’ learning trails anywhere, anytime (Song, 2014), and a lack of learning analytics tools to analyze, visualize and communicate the research findings (Siemens & Long, 2011). In addition, few studies adopted multiple methods to examine and triangulate the results of interactions.

Domination of assessment results after mCSCL practices

Almost half of the measurements were administered after the mCSCL activity. This implies that the results of the measurement could show only students’ individual products or outcomes within groups instead of providing a picture of the overall structure of flow of the group communication, or how individuals contribute to this process (Hmelo-Silver & Bromme, 2007). Such issues exist in the CSCL research literature, but appear to be more noticeable in mCSCL studies due to the constant changing of learning contexts (Looi et al., 2012). Therefore, students were not able to witness their collaborative learning process, identify problems and be active agents to improve their learning.

Short interventions and small sample sizes

Almost half of the studies did the measurement after the mCSCL practices and reported the improvement of collaborative performance and positive attitudes towards the collaboration. However, it is observed that many of these studies had short interventions with a small sample size. This triggers a doubt as to whether the positive results reported after mCSCL practices were due to the “novelty effect” (Thornton & Houser, 2005, p. 224), or the “Hawthorne effect” (Swan et al., 2005, p. 110); and whether the results of the measurement were significant with such a small sample size.

Lack of replication and sustainability of mCSCL research

Because many of these mCSCL studies tended to fall into the category of trials and pilots (Looi et al., 2012) without fine-grained details for measurement and without multiple types of methods, it is hard for these studies to be replicated in future mCSCL research, hence, it is even harder to be sustained.

Domination of in-class planned learning environments in the research design

In all the 35 eligible studies, 85% of them were conducted in “planned learning in-class” (65%) or “planned learning out-of-class” (20%); and 63% of all studies used the mobile devices as a collaborative learning system or tool in the collaborative process. This suggests that students’ learning processes were driven by the learning system or tool use in learning environments designed by the teacher rather than in authentic learning environments where students have the control of their own learning (Song, 2007).

Lack of the “mobility” nature of mCSCL practices

The most noticeable phenomenon found in this review is that in all the reviewed studies, the “mobility” nature of mCSCL practices was not adequately addressed from the lens of seamless learning (Wong & Looi, 2011). According to Looi et al. (2012), mCSCL does not simply mean “mobile + CSCL,” it indicates the changing practices that “mobile” technologies have initiated via continually re-constructed contexts and the instantaneous nature of
collaboration (Patten, Sánchez, & Tangney, 2006). This, in turn, implies that opportunities for student immediate mCSCL would make more knowledge generation possible, and further encourage active participation in the learning activity (Ryu & Parsons, 2012). However, the findings of this review show that in many of the eligible studies, the context of student learning has tended to be confined to “planned learning in-class” such as fixed physical classrooms, or “planned learning out-of-class” such as field trip types of studies (e.g., So et al., 2012). Few studies have attempted to measure collaborative learning across individual and social, physical and virtual and formal and informal learning simultaneously. The dynamic nature of mCSCL practices across different spaces remains scant although such practices have drawn researchers’ attention, and work has begun towards such endeavors (e.g., Arrigo et al., 2013; Song, 2014; Wong et al., 2012). In addition, in the majority of the studies, the mobile devices were provided by schools or institutions, which hinders students’ own exploration of the device to support their learning and makes students feel lack of ownership of their learning (Song, 2014).

**Discussions**

Compared to the findings and commentary from CSCL research literature (e.g., Hmelo-Silver & Bromme, 2007; Gress et al., 2010) and other related mobile educational application literature (e.g., Arrigo et al., 2013; Hsu, Y.-C., & Ching, 2013; Looi et al., 2012; Wong & Looi, 2011), findings from this mCSCL review show three distinctive features. First, more and more recent mCSCL studies attempted to capture contextualized just-in-time artifacts created by mobile devices and examine collaborative learning in authentic learning environments (e.g., So et al., 2012; Ryu & Parsons, 2012; Wong et al., 2012). Secondly, although existing issues of measurement in CSCL studies have also been identified in mCSCL practices, it is noted that these issues appear even more evident in mCSCL practices, such as a lack of replication of the practices across different contexts and collaborative models, and a lack of instruments for examining the collaborative process and the like. For example, although 63% of the studies have made use of the mobile devices as collaborative learning systems/tools, they have been used in a variety of ways for different purposes in different learning contexts ranging from using the system/tool as a medium for group collaborative discussion between students in groups to change the classroom dynamic (e.g., Cortez et al., 2005; Zurita & Nussbaum, 2007), as a way to develop student reading skills in second language (L2) learning (Lan et al., 2009), to being a model of personalized collaborative ubiquitous learning to support students doing tasks and activities (El-Bishouty et al., 2010). In addition, the measures are quite varied (10 different measures in the 35 studies). Thus, it is harder to replicate the practices and more challenging to capture and analyze the collaborative processes. Secondly, the measurement timing in mCSCL practices focuses on measurement after the collaborative activities (45% of the studies), the sample size tends to be small (60% of studies have a sample size from 10-50 participants) with a short intervention (about half of the studies conducted their intervention from 1 to 4 weeks), and many studies tend to be small-scale trials (Looi et al., 2012). Thus, despite that the research findings of mCSCL studies show favorable results, the effectiveness of measurement needs further justifications. Last but not least, although planned learning practices in both in-and out-of-class and design-based research are on the rise to foster learners’ collaborative learning, new endeavors are still scant to address methodological issues in mCSCL practices. Thus, the review study brings to light the following potential directions for further mCSCL research:

- Focus on using multiple methods across different measurement timings (before, during and after the mCSCL activity) in the design of mCSCL research (Gress et al., 2010; Hmelo-Silver & Bromme, 2007);
- Provide clear coding framework and measures to assess the process of collaboration in fine-grained detail (Hmelo-Silver & Bromme, 2007; Stahl, 2006);
- Design research with longer interventions and larger sample sizes to make the research results more robust;
- Adopt a method to collect and analyze big data (e.g., learning analytics) across different contexts to make the students visualize their collaborative learning process and guide them in the learning process with the ultimate goal of optimizing their collaborative knowledge construction and developing collaborative skills (Ogata et al., 2014; in the meantime provide opportunities for teachers to identify problems for pedagogical decision making (Long & Siemens, 2011);
Design research in which students’ collaboration distributes in different spaces (e.g., formal and informal, and virtual and physical learning spaces) using their own mobile devices with existing applications rather than using designed learning systems or tools; and

Lay emphasis on investigating viable and novel methodological approaches that address how to capture students’ collaborative process and outcomes in the mobile, reconstructed contexts. For example, Hakkarainen (2009) proposes using mobile devices for contextually and repeatedly sampling students’ knowledge practices in their natural context to examine students’ intellectual and emotional processes at personal and collective levels related to their trialogical (object-related) knowledge-advancement efforts. By doing so, important pedagogical implications can be uncovered in mCSCL practices.

Design mCSCL activities that span across different spaces to bridge formal and informal learning with social network platforms to develop students’ lifelong learning and collaborative skills.

Conclusions

This study conducted a systematic review of the methodological approaches of 35 eligible mCSCL studies. An analysis framework was developed to address four research questions regarding what methods have been adopted, when the measurements were administered, whether the methods are effective and what methodological issues are discovered. The research finding shows seven methods were employed to evaluate the 10 types of measures, and the timing of measurement spread across before, during and after mCSCL practices. By grouping the methods with the timing of measurement premised on the analysis framework and contextualizing the reviewed studies, it was found that in many cases, the methods might not be able to measure what the studies intended to measure effectively. Seven issues of methodological approaches were pinpointed and discussed. It should be noted that the findings of this review are limited due to the relatively small number of examples of studies analyzed. A further attempt is envisaged to critically review such issues with more detailed description and analysis in mCSCL studies.

Acknowledgements

The study was funded by Hong Kong Institute of Education under MIT Small Research Grant (Ref. MIT/SRG09/13-14).

References


Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence

Zacharoula Papamitsiou* and Anastasios A. Economides
Interdepartmental Progr. of Postgraduate Studies in Information Systems, University of Macedonia, 156 Egnatia Avenue, Thessaloniki, 54636, Greece // papamits@uom.edu.gr // economid@uom.gr

*Corresponding author

ABSTRACT
This paper aims to provide the reader with a comprehensive background for understanding current knowledge on Learning Analytics (LA) and Educational Data Mining (EDM) and its impact on adaptive learning. It constitutes an overview of empirical evidence behind key objectives of the potential adoption of LA/EDM in generic educational strategic planning. We examined the literature on experimental case studies conducted in the domain during the past six years (2008-2013). Search terms identified 209 mature pieces of research work, but inclusion criteria limited the key studies to 40. We analyzed the research questions, methodology and findings of these published papers and categorized them accordingly. We used non-statistical methods to evaluate and interpret findings of the collected studies. The results have highlighted four distinct major directions of the LA/EDM empirical research. We discuss on the emerged added value of LA/EDM research and highlight the significance of further implications. Finally, we set our thoughts on possible uncharted key questions to investigate both from pedagogical and technical considerations.

Keywords
Adaptive learning, Educational data mining, Empirical evidence, Learning analytics, Systematic review

Introduction
The information overload, originating from the growing quantity of “Big Data” during the past decade, requires the introduction and integration of new processing approaches into everyday objects and activities (“ubiquitous and pervasive computing”) (Cook & Das, 2012; Kwon & Sim, 2013). Handling large amounts of data manually is prohibitive. Several computational methods have been proposed in the literature to do this analysis.

In commercial fields, business and organizations are deploying sophisticated analytic techniques to evaluate rich data sources, identify patterns within the data and exploit these patterns in decision making (Chaudhuri, Dayal & Narasayya, 2011). These techniques combine strategic planning procedures with informational technology instruments, summarized under the term “Business Intelligence” (Eckerson, 2006; Jourdan, Rainer & Marshall, 2008). They constitute a well-established process that allows for synthesizing “vast amount of data into powerful decision making capabilities” (Baker, 2007, p. 2).

Recently researchers and developers from the educational community started exploring the potential adoption of analogous techniques for gaining insight into online learners’ activities. Two areas under development oriented towards the inclusion and exploration of big data capabilities in education are Educational Data Mining (EDM) and Learning Analytics (LA) and their respective communities.

EDM is concerned with “developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist” (Romero & Ventura, 2013, p. 12). Respectively, LA is an area of research related to business intelligence, web analytics, academic analytics, action analytics and predictive analytics. According to the definitions introduced during the 1st International Conference on Learning Analytics and Knowledge (LAK), LA is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and environments in which it occurs” (https://tekri.athabascau.ca/analytics/).

Explaining the previous definitions, LA and EDM constitute an ecosystem of methods and techniques (in general procedures) that successively gather, process, report and act on machine-readable data on an ongoing basis in order to advance the educational environment and reflect on learning processes. In general, these procedures initially emphasize on measurement and data collection and preparation for processing during the learning activities. Next, they focus on further analysis, reporting of data and interpretation of results, targeting to inform and empower...
learners, instructors and organization about performance and goal achievement, and facilitate decision making accordingly.

Both communities share similar goals and focus where learning science and data-driven analytics intersect. However, they differ in their origins, techniques, fields of emphasis and types of discovery (Chatti et al., 2012; Romero & Ventura, 2013; Siemens & Baker, 2012). Romero and Ventura (2013) presented an up-to-date comprehensive overview of the current state in data mining in education. In their overview, the authors do not present research results as empirical evidence. Their focus targets on the objectives, methods, knowledge discovery processes and tools adopted in EDM research. Analogous attempts were presented by Ferguson (2012) and Bienkowski et al. (2012) in the state of LA in 2012 and in an issue brief respectively.

All of these previous studies claim that as far as it concerns the approach to gaining insights into learning processes, LA adopts a holistic framework, seeking to understand systems in their full complexity. On the other hand, EDM adopts a reductionistic viewpoint by analyzing individual components, seeking for new patterns in data and modifying respective algorithms. In other words, these two research areas are complementary and in order to capture the whole picture, someone should follow their traces alongside each other.

**Motivation and rationale of the study**

The motivation for this review derived from the fact that empirical evidence is required for theoretical frameworks to gain acceptance in the scientific community. A search in relevant literature did not reveal any review of empirical evidence of the added value of research in both domains. Consequently, there was a need to supply the audience with an accredited overview. This paper aims to fill that gap.

The value of any single study is derived from how it fits with and expands previous work, as well as from the study’s intrinsic properties. Thus, putting together all the unbiased, credible results from previous research would be a step towards understanding the whole picture and construct a map of our knowledge on the domain. In a sense, the rationale of our study was to manage the overwhelming amount of publications through a critical exploration, evaluation and synthesis of the previous empirical results that worth reflection.

This paper’s goal is to carry out a systematic review of empirical evidence in order to contribute towards:

- a complete documentation of the applied research approaches so far,
- a feasibility study that captures the strengths and weaknesses of research in the domain, and
- the identification of possible threats, and thus motivate the research community redefine or refine related questions or hypotheses for further research (opportunities).

**The research questions**

The following research questions need to be addressed, and are distinguished into primary (generalized: set to fulfill the goals of the review) and secondary (sub-objectives/specific: refine the primary - explanatory):

**RQ1 (Primary)-Research Objectives:** Which are the basic research objectives of LA/EDM so far (in terms of measurable metrics), and which methods do researchers follow to achieve these goals?

- **RQ1.1 (Secondary) - Efficacy of implementation:** What are the significant results from previous research that constitute empirical evidence regarding the impact of LA/EDM implementation?

- **RQ1.2 (Secondary) - Interpretation of the results:** What do these results indicate regarding the added value of this technology?

**RQ2 (Primary) - Future challenges:** Which other emerging research technologies should be explored through the LA/EDM viewpoint?
Research methodology

The followed methodology qualifies this article as a systematic qualitative review of empirical research results concerning LA/EDM (Okoli & Schabram, 2010).

In order to conduct the literature review we defined a review protocol, consisting of four discrete stages: a) searching the literature – data collection, b) reviewing and assessing the search results – selection of primary studies, c) analyzing, coding and synthesizing the results, and d) reporting the review.

During the first stage, our goal was to collect the appropriate studies. For that reason, we determined and accessed the article pool and declared the key search terminology. We extensively and iteratively searched international databases of authoritative academic resources and publishers, including Scopus, ERIC, Google Scholar, Science Direct, DBLP and ACM Digital Library. We also scanned International Journals and selected Conference Proceedings. The search terms included learning analytics, learning analytics tools, learning analytics case studies, educational data mining, knowledge discovery in education. The search process spanned from March 2013 to August 2013. The time frame of the search was bound within the last six years (2008-2013), in which emergence and adoption of LA/EDM has grown.

Due to the orientation of our work towards the practical implementation and exploitation of LA/EDM, at the end of the data collection stage, we explicitly determined the article inclusion/exclusion criteria (Table 1).

<table>
<thead>
<tr>
<th>Table 1. Inclusion/exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Include</strong></td>
</tr>
<tr>
<td>Articles published in Journals with Impact Factor</td>
</tr>
<tr>
<td>Full-length articles published in International Conference/Workshop Proceedings</td>
</tr>
<tr>
<td>Present quantitative results</td>
</tr>
<tr>
<td>Date from 2008 to 2013</td>
</tr>
</tbody>
</table>

The search procedure, and after deleting the duplicate records, yielded 209 results. 40 of them are published in International Journals and 169 articles were presented at International Conferences. Then we assessed the quality of the collected literature according to the following rigorous quantitative/qualitative rules:

- Number of citations
- Degree of underlying innovation (e.g., significant changes in techniques/equipment or software, as proposed by UNESCO: [http://www.uis.unesco.org/ScienceTechnology/Pages/st-data-collection-innovation.aspx](http://www.uis.unesco.org/ScienceTechnology/Pages/st-data-collection-innovation.aspx))
- In depth illustration of the followed methodology (e.g., clear settings, fully-explained experimental procedure, etc.)
- Sufficient presentation of the findings (e.g., analytical discussion of findings and interpretation of results, use of figures and tables when needed, etc.)

Throughout the assessment procedure we identified that among the 209 retrieved articles, 40 of them were considered more central to our review (key studies), based on the combination of the above rules.

Next, we proceeded on with an article classification according to the adopted research strategy (category), research discipline (topic), learning settings, research objectives (goals), data gathering (sources and data-types) and analysis technique (method) and results. Finally, we used non-statistical methods to evaluate and interpret findings of the collected studies, and conduct the synthesis of this review.

Limitations

Although there are papers concerning EDM that are published before 2008, this year was a landmark for the independent growth of this research domain ([1st Conference in EDM - http://www.educationaldatamining.org/EDM2008/](http://www.educationaldatamining.org/EDM2008/)). For that reason we decided to examine the literature on
experimental case studies conducted in the domain from 2008 to 2013. Furthermore, indicative reviews on former work can be found in Romero and Ventura (2007) and Romero and Ventura (2010).

We should note that, despite the fact that appreciated research works have been published in respectable conferences like Intelligent Tutoring Systems (ITS), Artificial Intelligence in EDucation (AIED) and User Modeling, Adaptation and Personalization (UMAP), in this review we included published papers only from the EDM and LAK conferences. We acknowledge the leadenness of the previously mentioned conferences, but, in this work, we wanted to isolate the LA/EDM research and focus on its strengths, weaknesses, opportunities and threats.

We should also mention that the papers presented at the 6th International Conference on Educational Data Mining were excluded from the review process, since at the time of the search process these articles had had no citations. However, we decided to include recently published Journal articles (published in 2013) with lower number of citations, as indicative of current trends in the domain.

Results

In this section, we present our findings based on the analysis of the published case studies. We used non-statistical methods to evaluate and interpret findings of the collected studies.

According to the followed research strategy, most of the published case studies are exploratory or experimental studies. Some of them are evaluation studies, while others are empirical studies or surveys. Furthermore, the research topics differ from study to study, but most of them focus on science, technology, engineering, and mathematics (STEM).

Based on the learning settings of the studies (illustrated in Table 2), most studies are conducted within Virtual Learning Environments (VLEs) and/or Learning Management Systems (LMSs). Other popular learning settings are Cognitive Tutors (CTs), computer-based and web-based environments, mobile settings, and more recently, Massive Open Online Courses (MOOCs) and social learning platforms.

<table>
<thead>
<tr>
<th>Learning setting</th>
<th>Authors &amp; Year (Paper Ref.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLEs / LMSs†</td>
<td>Lin, Hsieh &amp; Chuang, 2009; Lykourentzou et al., 2009a; Lykourentzou et al., 2009b; Macfadyen &amp; Dawson, 2010; Merceron &amp; Yacef, 2008; Romero et al., 2008; Romero-Zaldivar et al., 2012; Tanes et al., 2011</td>
</tr>
<tr>
<td>MOOC/social learningb</td>
<td>Clow &amp; Makriyiannis, 2011; Fournier et al., 2011; Kizilcec et al., 2013</td>
</tr>
<tr>
<td>Web-based educationc</td>
<td>Abdous, He &amp; Yen, 2012; Giesbers et al., 2013; He, 2013; Khribi et al., 2009; Li et al., 2011; Romero et al., 2009</td>
</tr>
<tr>
<td>Cognitive tutorsd</td>
<td>Baker et al., 2008; Moridis &amp; Economides, 2009; Pardos et al., 2013; Shih, Koedinger &amp; Scheines, 2008</td>
</tr>
<tr>
<td>Computer-based educationf</td>
<td>Ali et al., 2012; Barla et al., 2010; Blikstein, 2011; Jeong &amp; Biswas, 2008; Levy &amp; Wilensky, 2011; Santos et al., 2012; Thai-Nghe et al., 2011</td>
</tr>
<tr>
<td>Multimodalityf</td>
<td>Worsley &amp; Blikstein, 2013</td>
</tr>
<tr>
<td>Mobilityg</td>
<td>Chen &amp; Chen, 2009; Leong et al., 2012</td>
</tr>
</tbody>
</table>

Note. †VLEs/LMSs: controlled environment (VLE/LMS), used for gathering learner and activity data, ‡MOOC/social learning: informal, social learning setting, §Web-based education: web-based e-learning environments except from VLEs, LMSs and MOOCs, †Cognitive tutors: special software, utilized for the needs of the study, ‡Computer-based education: other environments that include some type of computer technology (e.g. desktop applications, etc.) except from those belonging to one of the other categories, §Multimodality: learner data in different modalities, gMobility: mobile devices used as the primary learning mediator.

The authors gathered data from different data sources, including log files from the goal-oriented implemented systems, questionnaires, interviews, Google analytics, open datasets from the dataTEL Challenge (http://www.teleurope.eu/pg/groups/9405/datatel/), virtual machines, and many more. In particular, researchers tracked different types of data in order to measure students’ participation and login frequency, number of chat messages between participants and questions submitted to the instructors, response times on answering questions and...
solving tasks, resources accessed, previous grades, final grades in courses, detailed profiles, preferences from LMSs, forum and discussion posts, affect observations (e.g. bored, frustrated, confused, happy, etc.) and many more.

Another important parameter is the data mining method adopted by authors to analyze the gathered data. In the field of LA/EDM, the most popular method is classification, followed by clustering, regression (logistic/multiple) and more recently, discovery with models. In addition, algorithmic criteria computed for comparison of methods include precision, accuracy, sensitivity, coherence, fitness measures (e.g. cosine, confidence, lift, etc.), similarity weights, etc. Table 3 displays the classification of the key studies according to the data mining method they adopt.

**Table 3. Classification of case studies according to the analysis method**

<table>
<thead>
<tr>
<th>Data analysis method</th>
<th>Authors &amp; Year (Paper Ref.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>Baker et al., 2008; Barla et al., 2010; Chen &amp; Chen, 2009; Dejaeger et al., 2012; Dekker et al., 2009; Jeong &amp; Biswas, 2008; Guruler et al. 2010; Guo, 2010; Huang &amp; Fang, 2013; Khribi et al., 2009; Kizilec et al., 2013; Klašnić-Miličević et al., 2011; Li et al., 2011; Li et al., 2013; Lykourentzou et al., 2009a; Lykourentzou et al., 2009b; Moridis &amp; Economides, 2009; Pardos et al., 2013; Romero et al., 2008; Thai-Nghe et al., 2011</td>
</tr>
<tr>
<td>Clustering</td>
<td>Abdous, He &amp; Yen, 2012; Chen &amp; Chen, 2009; Khribi et al., 2009; Kizilec et al., 2013; Klašnić-Miličević et al., 2011; Lykourentzou et al., 2009b; Romero et al., 2009</td>
</tr>
<tr>
<td>Regression</td>
<td>Abdous, He &amp; Yen, 2012; Macfadyen &amp; Dawson, 2010; Romero-Zaldivar et al., 2012</td>
</tr>
<tr>
<td>Text mining</td>
<td>He, 2013; Leong et al., 2012; Lin, Hsieh &amp; Chuang, 2009</td>
</tr>
<tr>
<td>Association rule mining</td>
<td>Merceron &amp; Yacef, 2008; Romero et al., 2009</td>
</tr>
<tr>
<td>Social Network Analysis</td>
<td>Fournier et al., 2011; Macfadyen &amp; Dawson, 2010</td>
</tr>
<tr>
<td>Discovery with models</td>
<td>Ali et al., 2012; Pardos et al., 2013; Shih, Koedinger &amp; Scheines, 2008</td>
</tr>
<tr>
<td>Visualization</td>
<td>Clow &amp; Makriyiannis, 2011; Fournier et al., 2011; Santos et al., 2012</td>
</tr>
<tr>
<td>Statistics</td>
<td>Giesbers et al., 2013; Guo, 2010</td>
</tr>
</tbody>
</table>

The article classification according to the research objectives (goals) is illustrated in Table 4. As seen in this table, the majority of studies investigate issues related to student/student behavior modeling and prediction of performance, followed by increase of students’ and teachers’ reflection and awareness and improvement of provided feedback and assessment services.

**Table 4. Classification of case studies according to the research objectives**

<table>
<thead>
<tr>
<th>Research objectives (goals)</th>
<th>Authors &amp; Year (Paper Ref.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student/Student behavior modeling</td>
<td>Abdous, He &amp; Yen, 2012; Baker et al., 2008; Blikstein, 2011; Fournier et al., 2011; He, 2013; Jeong &amp; Biswas, 2008; Kizilec et al., 2013; Levy &amp; Wilensky, 2011; Li et al., 2011; Pardos et al., 2013; Romero et al., 2008; Shih, Koedinger &amp; Scheines, 2008</td>
</tr>
<tr>
<td>Prediction of performance</td>
<td>Abdous, He &amp; Yen, 2012; Huang &amp; Fang, 2013; Lykourentzou et al., 2009b; Macfadyen &amp; Dawson, 2010; Moridis &amp; Economides, 2009; Pardos et al., 2013; Romero et al., 2008; Romero-Zaldivar et al., 2012; Shih, Koedinger &amp; Scheines, 2008; Thai-Nghe et al., 2011</td>
</tr>
<tr>
<td>Increase (self-) reflection &amp; (self-) awareness</td>
<td>Ali et al., 2012; Clow and Makriyiannis, 2011; Fournier et al., 2011; Macfadyen &amp; Dawson, 2010; Santos et al., 2012</td>
</tr>
<tr>
<td>Prediction of dropout &amp; retention</td>
<td>Dejaeger et al., 2012; Dekker et al., 2009; Giesbers et al., 2013; Guo, 2010; Guruler et al. 2010; Kizilec et al., 2013; Lykourentzou et al., 2009a</td>
</tr>
<tr>
<td>Improve assessment &amp; feedback services</td>
<td>Ali et al., 2012; Barla et al., 2010; Chen &amp; Chen, 2009; Leong et al., 2012; Tanes et al., 2011; Worsley &amp; Blikstein, 2013; Wilson et al., 2011</td>
</tr>
<tr>
<td>Recommendation of resources</td>
<td>Khribi et al., 2009; Klašnić-Miličević et al., 2011; Romero et al., 2009; Thai-Nghe et al., 2011; Verbert et al., 2011</td>
</tr>
</tbody>
</table>
Key studies analysis

In this section we present the findings of the review process and answer on the initially set research questions RQ1 and RQ1.1. The rest of the research questions (mostly the results of the case studies and their comparative evaluation, as well as current and future trends, possible gaps and new research directions) are discussed in next section.

RQ1: Which are the basic research objectives of LA/EDM so far (in terms of measurable metrics), and which methods do researchers follow to achieve these goals?

Student/student behavior modeling

As seen from table 4, detection, identification and modeling of students’ learning behavior is a primary research objective. More specifically, the authors seek to identify learning strategies and when they occur, and model affective and metacognitive states (Abdous, He & Yen, 2012; Baker et al., 2008; Blikstein, 2011; Jeong & Biswas, 2008; Levy & Wilensky, 2011; Shih, Koedinger & Scheines, 2008). For example, Abdous, He and Yen (2012) and He (2013) tried to correlate interactions within a Live Video Streaming (LVS) environment to students’ final grades in order to predict their performance, discover behavior patterns in LVSs that lead to increased performance, and understand the ways students are engaged into online activities. In another case study, Blikstein (2011) logged automatically-generated data during programming activity in order to understand students’ trajectories and detect programming strategies within Open-Ended Learning Environments (OELEs). Furthermore, Shih, Koedinger and Scheines, (2008) used worked examples and logged response times to model the students’ time-spent in terms of “thinking about a hint” and “reflecting on a hint” for capturing behaviors that are related to reasoning and self-explanation during requesting hints within a CT environment. In another self-reasoning example, Jeong and Biswas (2008) tried to analyze students’ behavior based on the sequence of actions, and to infer learning strategies within a teachable agent environment.

Another orientation is the discovery and modeling of the respective behaviors within MOOCs (Fournier et al., 2011; Kizilcec et al., 2013). The authors tried to identify meaningful, high-level patterns of participation, engagement and disengagement in learning activities in this recently introduced learning setting.

Updated and extended review on student modeling approaches can be found in Chrysafiadi and Virvou (2013) and in Pena-Ayala (2014).

Prediction of performance

Authors also explore, identify and evaluate various factors as indicators of performance for prediction purposes. Among these factors, demographic characteristics, grades (in pre-requisite courses, during assessment quizzes and their final scores), students’ portfolios, multimodal skills, students’ participation, enrollment and engagement in activity and students’ mood and affective states are acknowledged as the most common ones (Abdous, He & Yen, 2012; Huang & Fang, 2013; Lykourentzou et al., 2009b; Macfadyen & Dawson, 2010; Moridis & Economides, 2009; Pardos et al., 2013; Romero-Zaldivar et al., 2012). For example, Macfadyen and Dawson (2010) examined the effect of variables tracked within an LMS-supported course (e.g., total number of discussion messages posted, total time online, number of web links visited, etc.) on students’ final grade. In another example, Lykourentzou et al., (2009b) used neural networks to accurately cluster students at early stages of a multiple choice quiz activity.

Moreover, researchers investigated affective factors that influence learning outcomes (within an ITS) (Pardos et al., 2013) and used simulation environments (virtual appliances that appear to learners as regular desktop applications) to monitor students and predict their performance (Romero-Zaldivar et al., 2012). Pardos et al. (2013) employed discovery with models, post-hoc analysis of tutor logged data and sensor-free detectors of affect (based on classification algorithms), while Romero-Zaldivar et al. (2012) tracked events (such as work-time, commands, compile, etc.) and analyzed the gathered data with multiple regression for the estimation of the variance of performance.
Increase (self-)reflection and (self-)awareness

Another crucial issue in EDM/LA research that authors attempt to address is how to increase the instructors’ awareness, identify “disconnected” students and evaluate visualizations regarding their capabilities on informing students about their progress and compared to peers. In order to provide instructors with pedagogically meaningful information and to help them extract such information on their own, the researchers embedded multiple representations of feedback types (Ali et al., 2012) and multiple widget technology for personalization of learning environments (Santos et al., 2012). Alternatively, content analysis for threaded discussion forums was explored regarding its monitoring capabilities (Lin et al., 2013). In particular, the authors aimed to facilitate the automated coding process within a repository of postings in an online course, in order less monitoring of the discussion to be needed by the instructor. Furthermore, Merceron and Yacef (2008) employed association rule mining to extract meaningful association rules to inform teachers about usage of extra learning material. The authors investigated students’ usage of learning resources and self-evaluation exercises and its possible impact on final grades.

In the context of social/open learning, the researchers explored the usefulness and motivation capabilities of dashboard-like applications regarding their self-reflection and self-awareness opportunities (Clow & Makriyiannis, 2011). In particular, the request was related to the effect of “expert users” presence on participants’ awareness of their own contribution and participation in an online reputation system with positive feedback only. Furthermore, Fournier et al. (2011) searched for crucial moments of learning based on interactions in MOOCs. In this case, the authors examined the impact of visualized provision of useful information to social learners regarding their participation and social interactions.

Prediction of dropout and retention

Prediction of dropout and retention are also key issues for LA/EDM research. In order to predict students’ dropout at early stages, Lykourentzou et al. (2009a) applied a combination of three machine learning techniques on detailed students’ profiles from an LMS environment. The authors compared the accuracy, sensitivity and precision measures of the proposed method to others in literature. From a similar point of view, Dekker et al. (2009) tried to predict students’ dropout and identify factors of success based on the use of different classification algorithms. The authors compared the accuracy and performance of these algorithms to make a selection between classifiers. In particular, they used classifiers for prediction of dropout based on simple “early” data (from first year enrollment) and boosted accuracy with cost-sensitive learning.

More recently, Kizilcec et al. (2013) classified learners according to their interactions (video lectures and assessment) with course content in learning activities in MOOCs. Next, they clustered engagement patterns, and finally, they compared clusters based on learners’ characteristics and behavior.

The issue of motivating engagement in learning activities and consequently increasing students’ satisfaction and retention was also explored (Dejaeger et al., 2012; Giesbers et al., 2013; Guo, 2010; Guruler et al., 2010). Demographics and factors like achievement rates and final performance were associated to students’ motivation to remain engaged and actively enrolled in courses. Identification of success factors urged Giesbers et al. (2013) to investigate the relationship between observed student behavior (i.e., actual usage of synchronous tools), motivation, and performance on a final exam. The researchers explored whether actual usage of synchronous tools increases the motivation to participate in online courses that support these tools. Similarly, Guo (2010) used statistical measures and neural network techniques for prediction of students’ retention. The researcher examined the number of students enrolled in each course and the distinction rate in final grades. Furthermore, Dejaeger et al. (2012) explored measures of students’ satisfaction for retaining student population. The authors investigated a number of constructs of satisfaction (e.g., perceived usefulness of training, perceived training efficiency, etc.) along with class related variables.

Improve feedback and assessment services

Many researchers have explored the use of LA/EDM in producing meaningful feedback. Feedback is strongly related to reflection and awareness and could be informative regarding students’ dropout intentions. For that reason, provision of appropriate forms/types of feedback was a major issue for Ali et al. (2012), Clow and Makriyiannis (2011) and Macfadyen and Dawson (2010), formerly presented. Visualization of feedback was also crucial for Tanes et al. (2011). The authors explored instructors’ perceptions of feedback types in relation to students’ success.
Complementary to that, in the mobile learning context, Leong et al. (2012) explored the impact and usefulness of SMS free-text feedback to teacher regarding the feelings of students, after a lecture. Their goal was to visualize positive and negative aspects of the lecture by taking advantage of the limited SMS length and the use of emoticons in order to provide free-text feedback to teacher.

In addition to these studies, an extensive area of LA/EDM research deals with issues related to using LA/EDM for adaptive assessment of goal achievement during activities. The landscape in this domain is quite distributed and diverse. Selection of the most appropriate next task during adaptive testing, students’ satisfaction level during mobile formative assessment, as well as construction of sophisticated measures of assessment (Barla et al., 2010; Chen and Chen, 2009; Wilson et al., 2011; Worsley & Blikstein, 2013) have emerged. Barla et al. (2010) focused on assessment capabilities of EDM methods and combined three different classification methods for selection of the most appropriate next task during adaptive testing. In a different context, Chen and Chen (2009) developed a tool that uses six computational intelligence theories according to the web-based learning portfolios of an individual learner, in order to measure students’ satisfaction during mobile formative assessment. Furthermore, Worsley and Blikstein (2013) aimed to detect metrics that could be used primarily as formative assessment tools of sophisticated learning skills acquisition in process-oriented assessment. A combination of speech recognition with knowledge tracing was proposed by the authors as method for multimodal assessment.

### Recommendation of resources

Another major issue in dataset-driven research concerns data resources and their management. Research in this domain focuses on a technical aspect. The approaches include similarity calculation mechanisms deployment, comparison of the performance of different mining algorithms, aggregation of different datasets in the context of dataset-driven research, suggestion of infrastructures for storing and forwarding learning-resources metadata (Romero et al., 2009; Thai-Nghe et al., 2011; Verbert et al., 2011) for resource recommendation in larger scale and across different contexts. Examples of algorithmic approaches also include recommendations according to the affective state of the learner (Santos & Boticario, 2012), implementation of collaborative filtering to sequence learning activities, hybrid recommendations based on learner and content modeling (Khribi et al., 2009; Klašnja-Miličević et al., 2011) and more.

Verbert et al. (2011) presented an analysis of publicly available datasets for TEL that can be used for LA in order to support recommendations (of resources or activities) for learning. In addition, the authors evaluated the performance of user-based and item-based collaborative filtering algorithms and measured their accuracy and coverage through metrics implementation. Moreover, Romero et al. (2009) explored user profile information and web-usage mining for recommendation of resources (here, hyperlinks). The authors compared the performance of three different mining algorithms.

A comprehensive review on recommender systems in the TEL context can be found in Manouselis et al. (2013).

### RQ1.1: What are the significant results from previous research that constitute empirical evidence regarding the impact of LA/EDM implementation?

According to the research objectives explored by the authors, Table 5a displays a categorization of the algorithmic-oriented findings from the collected studies.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Results</th>
</tr>
</thead>
</table>
| Student/ student behavior modeling | - Quantitative analysis could be applied for reporting on participants’ activity, while qualitative analysis could be more effective on revealing deeper concepts related to learning (Fournier et al., 2011).  
- Comprehensibility of the results strongly depends on human judgment - produced models are not equally interpretable by the teachers (Fournier et al., 2011; Romero et al., 2008).  
- Adding more predictor variables does not help improve the average prediction accuracy of the mathematical models explored by Huang and Fang (2013) for prediction of performance. However, neural networks method leads to better prediction results compared to those of the regression analysis method (Lykourentzou et al., 2009b). |
| Prediction of performance   |                                                                                                                                                                                                           |
Reducing the size of the training set by removing very high and very low probabilities of obtaining a correct answer without knowing the skill and obtaining an incorrect answer even though the student knows the skill, and forecasting techniques that embed sequential information (temporality) into the factorization process may improve the predictive model of students’ performance (Baker et al., 2008; Thai-Nghe et al., 2011).

Genre classification methods can automate the coding process in a forum and handle issues like imbalanced distribution of discussion postings (Lin, Hsieh & Chuang, 2009).

LMS log data are not data mining “friendly” (i.e., not stored the same way, data consolidation requires complex manipulations, etc.) (Merceron & Yacef, 2008).

Comparison of measures of interestingness of association rules did not significantly improved decision making for discarding a rule (Merceron & Yacef, 2008).

Combination of machine learning techniques afforded more reliable results, which depend on the level of detail of available students’ data (Lykourentzou et al., 2009a).

Simple classifiers had higher accuracy than sophisticated ones and cost-sensitive learning helps to bias classification errors (Dekker et al., 2009).

While investigating disengagement in MOOCs, the cross-cluster comparison can help understanding the reasons why learners remain to a cluster (Kizilcec et al., 2013).

A combination of students’ clustering and sequential pattern mining is suitable for the discovery of personalized recommendations (Romero et al., 2009), while content based filtering and collaborative filtering approaches are valid recommendation strategies (Khribi et al., 2009), but further research should be conducted (Verbert et al., 2011).

Table 5b displays a categorization of the pedagogy-oriented findings. The learning context of the studies has been taken under consideration, as well. That is because we wanted to maintain the targeted applicability of the results.

Table 6b. Classification of the results of LA/EDM case studies (pedagogical)

<table>
<thead>
<tr>
<th>Objective</th>
<th>Formal Learning</th>
<th>Non-Formal Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student/ student behavior modeling</td>
<td>Students’ detected critical moments during programming reflect students’ behavior and their perceived learning benefits, both in Secondary and Higher Education (Blikstein, 2011; Levy &amp; Wilensky, 2011). In secondary education, learning by teaching provides better opportunities for retaining metacognitive learning strategies (Jeong &amp; Biswas, 2008), while worked examples are effective indicators of self-explanation and learning gain (Shih, Koedinger &amp; Scheines, 2008). Specifying the moments that teacher should intervene requires to better distinguish between students who use worked examples, how they use them and their response times (Shih, Koedinger &amp; Scheines, 2008).</td>
<td>The presence of “experts” (that is users or organizations with advanced expertise or reputation on the field of study) has a significant impact on the highly unequal distribution of activities within a functioning social network (Clow &amp; Makriyannis, 2011). At a level of interactivity among learners or between learners and teachers, questions posed to instructors and chat messages posted among students (both in number and their content) are correlated (Abdous, He &amp; Yen, 2012). The discovery of four trajectories (auditing, completing, disengaging, sampling learners) (Kizilcec et al., 2013) roughly describes engagement that makes sense in MOOCs. Abdous, He and Yen (2012) couldn’t predict performance based on students’ participation and online interactions. Giesbers et al. (2013) and Macfadyen and Dawson (2010) found a significant positive relationship between participation and grades.</td>
</tr>
<tr>
<td>Prediction of performance</td>
<td>Both in Secondary and Higher Education, the number of quizzes passed is the main determinant of performance (i.e., the final grade), while others, such as number of posts, frequencies of the events and time-spent could identify activities that are related to higher or lower marks (Romero et al., 2008; Romero-Zaldívar et al., 2012; Shih, Koedinger &amp; Scheines, 2008).</td>
<td></td>
</tr>
</tbody>
</table>
In Secondary Education, engaged concentration and frustration are correlated with positive learning outcomes, while boredom and confusion are negatively correlated with performance (Pardos et al., 2013).

In Secondary Education, and from the instructors’ perspective, coding discussion posts in a forum can assist the teacher to automatically monitor the forum and maintain its quality (Lin, Hsieh & Chuang, 2009).

In Higher Education, meaningful rules increases teachers’ awareness regarding the students’ usage of additional material within LMSs (Merceron & Yacef, 2008).

SMS text increases instructor’s awareness on students’ affective states in order to modify the lecture (Leong et al., 2012).

Dashboard-like applications and multiple feedback representations could increase (self-)awareness and perceived value of provided feedback (Ali et al., 2012; Macfadyen & Dawson, 2010; Santos et al., 2012).

From the learners’ point of view, students want to be aware of what their peers are doing, but they don’t like to be tracked outside a course environment due to privacy concerns (Santos et al., 2012).

Identification of disconnected students based on their networking activity ended up to clusters of students with similar participatory behavior (Macfadyen & Dawson, 2010).

In MOOCs, the most common detected reasons for disengagement were personal commitments, work conflict and course overload (Kizilcec et al., 2013).

In web-videoconference settings there was not found a relation between motivation to participate and dropout (Giesbers et al., 2013).

Types of registration to the university as well as the family income seem to affect more the students’ retention (Guruler et al., 2010).

In Higher Education, adaptive selection of the most appropriate next task improved testing outcomes mostly for below-average students (Barla et al., 2010).

In Elementary Education, web-based learning portfolios of an individual learner during mobile formative assessment granted similar results to those of summative assessment (Chen & Chen, 2009).

Additional learner attributes (e.g., experience level indicators, learning interests, learning styles, learning goals and competences and background information), student’s expected performance on tasks, his recent navigation history (within a number or resources) or learner’s affective traits should be taken under consideration in recommendation processes (Khribi et al., 2009; Klašnjka-Milićević et al., 2011; Santos & Boticario, 2012; Thai-Nghe et al., 2011; Verbert et al., 2011).

Discussion and future research

From the former analysis it becomes apparent that recently, the educational research community has started applying sophisticated algorithmic methods on gathered (mostly raw) data for understanding learning mechanisms through an in-depth exploration of their relations and meaning. As seen in Tables 5a and 5b, the landscape of the LA/EDM research combines diverse and often conflicting aspects and results related to gaining insight into learning processes. However, the above results have highlighted four distinct major axis of the LA/EDM empirical research including:
- **Pedagogy-oriented issues** (e.g., student modeling, prediction of performance, assessment and feedback, reflection and awareness): several studies focus on pedagogically meaningful analysis on collected students' data in order to shed light to the whole picture from students/students' behavior modeling to self-regulated learning.

- **Contextualization of learning** (e.g., multimodality, mobility, etc.): a number of studies gathered data from the learning context itself and focus on positioning learning within specific conditions and attributes.

- **Networked learning** (e.g., MOOCs, social learning platforms, etc.): some case studies try to identify patterns within the social aspect of learning and the MOOCs, where the number of participants rapidly increases and the interactions between learners and the learners and the content are text/video-based.

- **Educational resources handling**: fewer, but not neglected studies raise the issue of organizing and recommending educational resources from data pools, and selecting the most appropriate algorithmic method for making suggestions.

However, these four axis are not completely autonomous, since significant overlaps may occur. For example, student modeling (i.e. a pedagogy-oriented issue) can still be explored in MOOCs (i.e., a form of Networked learning). However, this statement could only constitute a limitation which does not deduce the added value of the findings.

**RQ1.2: What these results indicate regarding the added value of this technology?**

One of the most important goals of the systematic review was to reveal the added value of the field explored. From the above analysis of findings derives that analysis of user interactions in order to “control” the information generated through technology has always been a request. LA/EDM research results indicate that data integration from multiple sources can improve the accuracy of a learner profile and subsequent adaptation and personalization of content. Exploration of students’ behavior within educational contexts that support multimodality and mobility could lead to shaping a holistic picture of how, when and where learning occurs.

Researchers set the educational context within limits in which previously it was almost impossible to infer behavior patterns, due to their high levels of granularity. In such advanced learning contexts, LA/EDM research community determines simple and/or sophisticated factors as predictors of performance and explores their predictive value and capabilities by tracking actual data and changes on behavioral data. The goal is to identify the most significant factors in order to develop better systems. These systems will allow students to monitor their own progress and will help them evaluate and adjust their learning strategies to improve their performance in terms of learning outcomes.

Moreover, the social dimension of learning and the opportunity of selectively participating in MOOCs are also explored with encouraging results. Consequently, the research community could gain insight into the learning mechanisms that previously were a “black box.”

**RQ2: Which other emerging research technologies should be explored through the LA/EDM viewpoint?**

Complementary, the literature overview has revealed a number of unexplored issues in this rapidly grown domain, including (but not limited to) the following:

**Suggested incorporation of other emerging research technologies with LA/EDM**

Game-based learning (GBL) has been acknowledged for its positive impact on learners. According to Collony et al. (2012, p.1), “playing computer games is linked to a variety of perceptual, cognitive, behavioral, affective and motivational impacts and outcomes.” One interesting research question is if and how LA/EDM methods could be applied to report and visualize learning processes during GBL. In other words, how can LA/EDM be applied on GBL to detect patterns and construct measures that are transferable to other OELs, in order to assess advanced skills development.
Another field evolving in a rapid pace is mobile and ubiquitous learning. Mobile learning has been acknowledged for the unique opportunity of offering authentic learning experiences anytime and anywhere (Tatar et al., 2003). Although two of the selected studies were conducted in a mobile context (Chen and Chen, 2009; Leong et al., 2012), none of them associated or explored the effect of the context on the attained results. LA/EDM research could investigate the appropriateness of the popular methods in the above context in order to provide sophisticated, personalized learning services through mobile applications.

Furthermore, according to Piaget’s theory of cognitive development, sensorimotor learning is the first stage of human learning (Piaget, 1952). Sensorimotor learning refers to improvement, through practice, in the performance of sensory-guided motor behavior (Krakauer and Mazzoni, 2011). Due to its high relevance to the brain anatomy and functionality, sensorimotor learning has recently been under the lenses of neuroscience research (e.g., Catmur, 2013). LA/EDM has not been previously examined for sensorimotor learning or combined to neuroscience research. It would be interesting to study transformation of learning experience into strategy development (knowledge transfer) by exploring big neuroscience data.

Technology acceptance is also a well addressed issue in educational research. Regarding learning analytics acceptance, Ali et al. (2012) proposed a model that considers only two parameters – ease of use and perceived usefulness. However, more parameters should be explored in order to create a reliable learning analytics acceptance model. An appreciated model for computer based assessment acceptance was proposed by Terzis and Economides (2011). Researchers from the LA/EDM domain could also examine respective models that are suitable for the purposes of LA tools.

Finally, the review process didn’t yield any article related to learning “meta”-analytics (i.e., feeding machine readable results from the LA/EDM procedures to another data-driven system for diving decision making without the mediation of the human judgment parameter). It would be interesting to take advantage of the plethora of results from LA/EDM research towards introducing innovative intelligent tutoring systems or fully automated educational recommender systems.

Conclusions

Previous literature reviews on LA/EDM research provided significant insight into the conceptual basis of this rapidly growing domain. However, these studies did not conduct an analysis of actual research results. The current paper presents a systematic review of empirical evidence of LA/EDM research. We searched the literature and gathered representative, mature and highly-cited articles of real case studies with actual data, both from LA and EDM domains. The analysis of selected case studies and their results shed light on the approaches followed by the respective research communities and revealed the potential of this emerging field of educational research. Along with the arising opportunities, we discovered a number of gaps that require the researchers’ attention. Table 6 illustrates our findings regarding the strengths, weaknesses, opportunities and threats (SWOT) of LA/EDM research.

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Large volumes of available educational data</td>
<td>• Misinterpretation of results due to human judgment factors - focus on reporting, not decision.</td>
</tr>
<tr>
<td></td>
<td>• Heterogeneous data sources: not yet a unified data descriptive vocabulary – data representation issues.</td>
</tr>
<tr>
<td></td>
<td>• Mostly quantitative research results. Qualitative methods have not yet provided significant results.</td>
</tr>
<tr>
<td>• Use of pre-existing powerful and valid algorithmic methods.</td>
<td>• Information overload – complex systems.</td>
</tr>
<tr>
<td>• Interpretable multiple visualizations to support learners/teachers.</td>
<td>• Uncertainty: “are we ready yet?” So far, only skilled teachers/instructors could interpret the results correctly.</td>
</tr>
<tr>
<td>• More precise user models for guiding adaptation and personalization of systems.</td>
<td></td>
</tr>
<tr>
<td>• Reveal critical moments and patterns of learning.</td>
<td></td>
</tr>
<tr>
<td>• Gain insight to learning strategies and behaviors.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Use of Open Linked Data for data standardization and compatibility among different tools and applications</td>
<td>• Ethical issues – data privacy.</td>
</tr>
<tr>
<td></td>
<td>• Over-analysis: the depth of analysis becomes profound and the results lack generality. The “over-granularity” approaches so far might threaten the holistic picture</td>
</tr>
<tr>
<td>• Multimodal and affective learning opportunities</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. SWOT of LA/EDM research
based on sophisticated metrics.
- Self-reflection/self-awareness/self-learning in intelligent, autonomous and massive systems.
- Feed machine readable results from the LA/EDM procedures to other data-driven systems for diving decision making.
- Acceptance Model: e.g., perceived usefulness, goal expectancy, perceived playfulness, trust, etc.

Beyond learning perceptions and attitudes collected through questionnaires, every “click” within an electronic learning environment may be valuable actual information that can be tracked and analyzed. Every simple or more complex action within such environments can be isolated, identified and classified through computational methods into meaningful patterns. Every type of interaction can be coded into behavioral schemes and decoded into interpretable guidance for decision making. This is the point where learning science, psychology, pedagogy and computer science intersect. The issue of understanding the deeper learning processes by deconstructing them into more simple, distinct mechanisms remains in the middle of this cross-path.

We believe that this active research area will continue contributing with valuable pieces of work towards the development of powerful and mostly accurate learning services both to learners and teachers.

References


63


Review of Speech-to-Text Recognition Technology for Enhancing Learning

Rustam Shadiev¹, Wu-Yuin Hwang², Nian-Shing Chen³ and Yueh-Min Huang¹∗

¹Department of Engineering Science, National Cheng Kung University, No.1, University Road, Tainan 70101, Taiwan // ²Graduate Institute of Network Learning Technology, National Central University, No.300, Jhongda Road, Jhongli 32001, Taiwan // ³Information Management Department, National Sun Yat-sen University, No.70, Lienhai Road, Kaohsiung 80424, Taiwan // rustamsh@gmail.com // wyhwang@cc.ncu.edu.tw // nschen@mis.nsysu.edu.tw // huang@mail.ncku.edu.tw

*Corresponding author

ABSTRACT
This paper reviewed literature from 1999 to 2014 inclusively on how Speech-to-Text Recognition (STR) technology has been applied to enhance learning. The first aim of this review is to understand how STR technology has been used to support learning over the past fifteen years, and the second is to analyze all research evidence to understand how Speech-to-Text Recognition technology can enhance learning. The findings are discussed from different perspectives as follows: (a) potentials of STR technology, (b) its use by specific groups of users in different domains, (c) quantitative and/or qualitative research methodology used, and (d) STR technology implications. Some STR literature review showed that in earlier stage of development, the STR technology was applied to assist learning only for specific users, i.e., students with cognitive and physical disabilities, or foreign students. Educators and researchers started to apply STR technology in a traditional learning environment to assist broader group of users, while STR technology has been rapidly advancing over the years. The review revealed a number of distinct advantages of using STR learning. That is, STR-generated texts can greatly help students understand learning content of a lecture better, to confirm missed or misheard parts of a speech, to take notes or complete homework, and to prepare for exams. Furthermore, some implications over STR technology in pedagogical and technological aspects were discussed in the review, such as the design of technology-based learning activities, accuracy rate of the STR process and learning behaviors to use STR-texts that may limit the STR educational value. Thus, the review further discussed some potential solutions for the future research.

Keywords
Speech-to-text recognition technology, Literature review, Supporting and enhancing learning, Group of users

Introduction

Recent evidence suggests that some challenges and limitations exist in physical and online synchronous learning environments that still require attention to solve them (Camiciottoli, 2005; Miller, 2007; Chen, Ko, Kinshuk, & Lin, 2005; Huang & Chiu, 2014; Neilson, 2009; Nisbet & Spooner, 1999; Shadiev, Hwang, & Huang, in press; Wang, Chen, & Levy, 2008). For example, on an academic event, information is usually addressed through audio channels so that students with learning or physical disabilities, foreign students, and other at risk populations are challenged to understand the content (Camiciottoli, 2005; Lee, 2011; Miller, 2007; Nisbet & Spooner, 1999). Furthermore, one of the most common concerns reported in relation to online learning literature is the poor audio quality due to restricted internet bandwidth availability and traffic congestion (Chen et al., 2005; Wang et al., 2008). These problems can hinder students’ understanding of a delivered speech, and this may hamper students from engaging in classroom participation and interaction (Camiciottoli, 2005; Miller, 2007; Chen et al., 2005; Wang et al., 2008).

According to related literature, abovementioned problems can be solved by adopting some assistive media-to-text recognition technologies, such as writing-to-text, image-to-text, diagram-to-text, text-to-speech, speech-to-text, and handwriting-to-text. For example, Speech-to-Text Recognition (STR) technology synchronously transcribes text streams from speech input and shows them on a whiteboard or students’ computer screens (Alapetite, Andersen, & Hertzum, 2009; Fichten et al., 2000; Hwang, Shadiev, Kuo, & Chen, 2012; Jones, 2005; Konur, 2007; Kuo, Shadiev, Hwang, & Chen, 2012; Shadiev, Hwang, & Huang, 2013). It is suggested that STR-generated texts can greatly help students attain a better understanding of a lecture, do simultaneous note-taking during lectures, and complete homework (Hwang et al., 2012; Kuo et al., 2012; Shadiev et al., 2013). Furthermore, it is argued that STR-generated text can be employed as an additional text confirmation of what is being said, and it aids comprehension in case when listeners are students with learning or physical disabilities, foreign students, and other at risk populations (Shadiev et al., 2013; Wald & Bain, 2008).
The pedagogical usefulness of STR-technology application to enhance students’ learning was emphasized in several studies. The following are a few examples. The Speech Recognition in Schools Project (Nisbet & Wilson, 2002; Nisbet, Wilson, & Aitken, 2005) helped students to overcome difficulties in reading, writing, and spelling. The project presented significant improvements in some students’ basic reading, writing, and spelling skills with the support of STR. Wald and Bain (2008) developed STR applications to assist deaf students and non-native speakers to be involved in lectures. According to their research, students perceived that text generated by STR could improve learning if its accuracy is fairly good (Colwell, Jelfs, & Mallett, 2005; Wald & Bain, 2008). Ryba et al. (2006) examined the application of STR in a university lecture theatre attended by students who were native and non-native speakers of English. A non-native English lecturer delivered a course about information system to the participants by using STR. The participants reported that the system was a potentially useful instructional support mechanism; however, a greater accuracy in lecture text vocabulary recognition of the system needs to be achieved. Shadiev et al. (2013) applied STR technology to assist non-native English participants to learn at a seminar in English. It was found that most participants perceived that transcripts were useful for learning. Moreover, nineteen learning strategies to use transcripts were discovered, and participants with different learning achievements demonstrated different learning behaviors to use transcripts. Hwang et al. (2012) and Kuo et al. (2012) employed the STR technology in teaching and learning activities in online synchronous learning environment. Compared to students who did not use transcripts, it was found that students who used transcripts showed improvement on homework accomplishments and post-test results (Hwang et al., 2012; Kuo et al., 2012).

This study aims to review previous STR technology relevant literature and how it can enhance learning. STR technology was mostly used to assist specific groups of students (i.e., students with learning or physical disabilities or foreign students) in order to guarantee them the equal access to learning. However, as time passed by, the target users involved into research on STR technology has got broader. That is, nowadays STR technology is adopted to assist not only students with special needs but also general population of students for more educational purposes, such as enhancing students’ understanding of a presented learning content during and after academic activities as well as offering students guidance to accomplish reflective writing and homework. Furthermore, due to recent improvement of STR technology, particularly its accuracy rate, the technology is also adopted to support collaborative learning activities with multiple participants speaking simultaneously, such as group discussions or students’ oral presentations. Therefore, this study particularly summarized STR development history and its usage by specific group of users. First, this study looks into how STR technology has been used in education over the past fifteen years by reviewing relevant research. Second, this study demonstrates how effective STR technology can be to enhance learning for different groups of users, such as students with learning or physical disabilities, foreign students, online students, and students who study in physical environment. This study further highlights findings on STR technology and proposes several suggestions for future research. The following research questions were addressed in this review:

- How has Speech-to-Text Recognition technology been used in learning over the past fifteen years?
- What learning activities can make the most of Speech-to-Text Recognition technology and bring out the best learning outcomes to enhance learning?

Methodology

The literature from 1999 to 2014 inclusively were searched using the search terms such as speech-to-text, voice-to-text, speech recognition, transcription, and learning from ACM Digital Library, EBSCO Discovery Service, ERIC, PsychINFO, and Social Sciences Citation Index databases. A total of 42 selected articles were reviewed. Primary data source for this review include peer-reviewed journal articles, conference proceedings, and frequently cited books. The references provide a complete list of all the articles reviewed for this project (marked with an asterisk). The publications reviewed are organized into four dimensions that address (a) potentials of STR technology, (b) its use by specific groups of users in different domains, (c) research findings from studies using quantitative and/or qualitative methodologies, and (d) issues and considerations of applying STR technology. These categories provide an organizational framework to understand how STR technology has been used in learning, and to explore any research evidence in terms of how Speech-to-Text Recognition technology can enhance learning.

Findings of this review were organized into two particular aspects. The first aspect is STR methodology and approach. That is, this review aimed to understand how the STR technology has been applied to support learning. Findings in this aspect are reported based on STR technological development. In earlier stages, STR technology was
not as well-developed as it is now. One major issue was how to generate a satisfactory accuracy rate of transcripts from a speaker’s speech. Therefore, earlier attempts were made to apply STR technology only for particular groups of users, such as student with cognitive or physical disabilities. Afterwards, a lot of studies of how computers can assist language learning were carried out with applications of STR technology. Finally, STR technology was developed more mature and reliable; the accuracy rate of recognition voice into text became higher and even STR could transcribe multiple speakers at the same time. Thus, some experts applied STR in a traditional classroom during lectures or collaborative learning activities on other fields of knowledge. The other aspect is potentials and findings of STR technology to facilitate learning. That is, this review attempted to analyze all research evidence that how Speech-to-Text Recognition technology can enhance learning. Findings in this aspect centered on applications of the STR technology to support learning of different groups of users in traditional and online learning environments.

The usage of the STR in learning

This section analyzes findings from other studies regarding how the STR technology was applied in learning in the past fifteen years. Findings in this section are classified into the following categories: the usage of STR to assist learning of students (1) with cognitive or physical disabilities, (2) online students, (3) non-native speakers, and students (4) in traditional learning environment, and (5) in collaborative academic activities. Main findings of this section are summarized in Appendix.

Students with cognitive or physical disabilities

According to Lee (2011), Neilsen (2009), Nisbet and Wilson (2002), Nisbet et al. (2005), and Zhili, Wanjie and Jiacheng (2010), many students, who need additional support, have difficulties in reading, writing, or spelling, due to motor difficulties, visual impairment, or specific learning difficulty. Elliot, Foster, and Stinson (2002) suggested that students with hearing impairments rely on either reading lips or watching an interpreter to access to what the instructor spoke. It is extremely difficult for these students to focus their visual attention on note-taking and the instructor (or interpreter) simultaneously. Therefore, it was suggested to apply assistive technologies, such as a speech-to-text support service, to enhance computer-assisted learning for students with different types of disabilities.

In the Speech Recognition in Schools Project (Nisbet & Wilson, 2002; Nisbet et al., 2005), STR technology was used by secondary school students with special educational needs for one semester. Forty students with reading, writing or spelling difficulties from different schools in Scotland participated in this project. The project provided students with Dragon Naturally Speaking or IBM ViaVoice speech to text recognition software, and technical support and training delivered on site. Besides, a training pack was designed by the project for students to learn how to use one or other kinds of software.

IBM ViaVoice software was applied to assist students with hearing impairments to listen to lectures in three Canadian high schools (Leitch, 2008) and one university in UK (Wald, 2010; Wald & Bain, 2008). In the studies of Leitch (2008), Li et al. (2011), Wald (2008), Wald (2010), and Wald and Bain (2008), teachers were engaged in training and testing the STR technology for a minimum period of two weeks. When accuracy rates reached a certain satisfactory level of at least more than 80%, teachers employed STR technology when giving lectures. During this time, teachers displayed text to students. Most students in the studies of Leitch (2008), Li et al. (2011), Wald (2010), and Wald and Bain (2008) used STR-generated texts as an additional resource to verify and clarify what has been said by the lecturer as well as to take their own notes and argue their own opinions.

Elliot et al. (2002) carried out another study with USA high school and college students who were deaf and hard of hearing. Students were provided with notes from a speech-to-text support service called C-Print. The purpose was to help students to fill in gaps in their understanding of what transpired during class. This STR system produced a real-time transcripts displayed for students on their personal or laptop computers to access the information. A C-Print notes include as much information as possible, generally providing almost all of the meanings of the spoken lecture content. After class, notes were saved and edited so students could use them in paper or electronic format. The study with high school students lasted for approximately ten weeks and with college students for the period between ten to sixteen weeks. Analyzing students’ learning behaviors, Elliot et al. (2002) found that high school students typically read the notes only, while college students used multiple study strategies with the notes.
Online students

Network traffic congestion can cause poor quality of audio communication in a synchronous cyber classroom. Under such condition, students are not able to hear a speaker clearly. This issue was viewed as one technological challenge. It negatively affects online teaching and learning activities as it hinders students’ understanding of a delivered speech, and it also hampers students from engaging in classroom participation and interaction (Chen et al., 2005; Hwang et al., 2012; Kuo et al., 2012; Wang et al., 2008). To address this issue, Hwang et al. (2010), Hwang et al. (2012), Kuo et al. (2011), Kuo et al. (2012), and Shadiev (2011) employed Windows Speech Recognition in the Microsoft Operating System for STR tools to support various teaching and learning activities (lectures, oral presentations, and discussions) and communication for students in a synchronous learning environment. This application in Microsoft Operation system is the most available tool to get for the students and teachers to participate in the experiment. Way et al. (2008) argued that this application is similar to other open-source products that are easy to use and functions well, and it is available at no additional cost for users. The teacher and students of these two studies were given a training session of STR for three to six weeks before using it for teaching and learning activities. They dictated two articles over their local STR, and the content of the articles are related to the target teaching and learning activities. Thus, the STR technology could “learn” a speaker’s voice and terminology for specific field during the training, and then achieve a certain level of accuracy rate when being applied in the activities. A speaker’s speech was transcribed by the STR technology into text which was displayed simultaneously to students on their computer screens. Thus, the students could listen to a speaker and read the transcripts at the same time. More importantly, STR-generated text was saved for further revision to fix some recognition errors, and the students could obtain a nearly verbatim transcript to study it after the activities and to complete summary writing tasks.

Non-native speakers

According to Camiciottoli (2005), and Wu and Alrabah (2009), non-native speaker students have difficulties listening to content of an academic speech in a language other than their mother tongue (Bennett, Hewitt, Mellor, & Lyon, 2007; Camiciottoli, 2005; Wu & Alrabah, 2009). Bain et al. (2005) defined this problem as “accessibility issue.” Many of non-native speaker students flounder in delivering speeches in a foreign language and they need to make extra efforts when attempting to comprehend these speeches (Miller, 2007). Studies conducted with non-native speaker students revealed that they are silent and, generally, they engage in limited classroom participation and interaction (Bain et al., 2005; Camiciottoli, 2005; Miller, 2007). For this reason, speakers who give a speech to a non-native speaking audience should be aware of the potential obstacles and consider the need of delivery method variation. Li et al. (2011) and Wald and Bain (2008) proposed application of the speech-to-text recognition technology as a potentially reliable tool for non-native speaker students to better understand a speech given in a foreign language.

Ryba et al. (2006) applied IBM’s ViaVoice and Viascribe STR technology in an university lecture theater with one hundred sixty participants. Half of the participants were native speakers of English while the other half were not. Three 2-hour lectures lasting over a three-week period were delivered to the participants by a non-English speaking lecturer; yet, the lectures were given in English. The lecturer was trained to use the STR system and then applied it during the lectures. Spoken lectures were transcribed into text through STR technology, and then the text was displayed on a large screen in front of the lecture theatre so that students could both see and hear the lecture. After the lecture, the STR-generated texts could be saved and edited, punctuations were inserted, recognition errors were corrected and redundancies removed. The STR-generated text was accessible via the internet.

Coniam (1999) carried out a study in which Dragon Naturally Speaking STR system was employed to assist students to learn English as a second language. A small group of very competent second language subjects participated in the study. First, the subjects had the system recognized their own voices during 45 minutes by reading a training text of 3800 words. Next, the subjects read a text into the voice recognition software and this voiced text was analyzed to compare with that generated from voice of another small group of native speakers.

Shadiev et al. (in press) and Shadiev et al. (2013) applied Windows Speech Recognition built in the Microsoft Operating System to an eight-week graduate seminar program on advanced learning technologies. Seminar in English was executed once a week for non-native English participants. In the program, the participants needed to give a speech that would be graded and the STR technology generated transcripts from the speech. In order to
achieve a good accuracy rate of the STR, every participant was trained to use the system beforehand. Besides, transcripts were edited to fix some recognition errors and nearly verbatim transcripts were projected on the whiteboard along with the Microsoft Power Point slides of a presenter. Furthermore, transcripts were available online so that participants could study them during and after the seminar.

In the other study, Shadiev and his colleagues (2014) employed Windows Speech Recognition built in Microsoft Operating System for two lectures in English to assist non-native English participants to better comprehend lectures’ content. The difficulty of the first lecture was intermediate level and the other one was advanced level. Both lectures were delivered to participants through computer screens. Participants could see video of the instructor, slides of the lectures, and text generated by the STR. Shadiev et al. (2014) investigated participants’ visual attention on STR-generated text by employing eye-tracking technique. How differently effective STR-texts can be to influence participants’ learning achievement was also assessed. Besides, Shadiev et al. (2014) compared visual attention and learning behaviors to different characteristics of participants, such as learning ability, learning style preferences, and gender, in using STR-texts. Finally, students’ perceptions regarding usefulness of STR-texts for learning were also explored.

In the study of Weggerle, Schmidt, and Schulthess (2009), the spoken lecture was recorded and automatically transcribed by using the Naturally Speaking Software. The quality of the transcription was improved through a subsequent editing and correction process, and the pages from the beamer presentation were added to the time line of the text. STR-generated lecture notes were available as an interactive web application; they were searchable and allowed for individual annotations. The lecture notes were available for students for one semester in Technical Informatics course. Weggerle and colleagues aimed to enhance the learning efficiency for foreign language students and for those who didn’t attend the course but allowed to access to the lecture notes.

**Students in traditional learning environment**

Luppi et al. (2009) argued that the adoption of the STR technology in traditional learning environment has several benefits. One of them is to improve teaching methods and to enhance learning opportunities. For example, by using the STR, teachers can take a proactive, rather than a reactive approach to teach students with different learning styles. It provides educators with a practical means of making their teaching accessible and improves the quality of instruction in the process (Luppi et al., 2009).

Ranchal et al. (2013) adopted IBM ViaScribe and IBM Hosted Transcription Service to assist university students in the lectures of life and social sciences courses. Two distinct methods of the STR-mediated lecture acquisition, such as real-time captioning and post lecture transcription, were evaluated in the study. The instructor underwent initial voice training to develop a voice profile for the systems and to improve STR accuracy before starting real-time captioning or post-lecture transcription. During class, the STR processed verbal information into textual captions and streamed them on a screen or students’ computers. Students received drafts, unedited transcripts during lectures. However, errors in the transcribed text were then corrected and it was available for students after class.

Ryba et al. (2006) applied IBM ViaVoice and Viascribe systems to listen to lectures about information systems in an university lecture theatre with more than one hundred students. First, the lecturer underwent training to develop a voice profile for the system and to achieve a high level of accuracy through inputting dialogue and vocabulary into the system. Then, the STR systems transcribed lectures into texts which were displayed on a large screen in front of the lecture theatre. Besides, the edited STR-generated texts were delivered to students after the lecture via the internet. Ryba et al. (2006) explored students’ perceptions of using STR-texts and to what extent students make use of them. Moreover, the main advantages and limitations of using STR-texts were investigated.

Goddard, Kaplan, Kuehnle, and Beglau (2007) applied Read & Write GOLD system as cognitive prosthesis to support students’ learning needs in general education classrooms. Teachers and students were trained to use the system first. Then students used STR to reach certain level of understanding of learning material that they couldn’t understand before. As soon as students caught up the level of their peers’ understanding, they reduced or abandoned usage of the STR.
The STR technology applications used for supporting collaborative learning activities

According to Wald and Bain (2008), the STR usage to date focus primarily on situations where there is only one speaker, i.e., one-way lecturing. They argued this is a limited scenario because many occasions involve multiple speakers. Hwang et al. (2012), Kheir and Way (2006), Li et al. (2011), Wald and Bain (2008), and Zschorn, Littelfield, Broughton, Dwyer, and Hashemi-Sakhtsari (2003) suggested that STR technology can be used for multiple speakers environment as well; it may aid in producing transcripts from discussions, meetings, and other collaborative activities.

The STR applications were developed in studies of Fiscus, Ajot, and Garofolo (2007), Wald and Bain (2008), and Zschorn et al. (2003) for multiple-speakers environment.

In the study of Wald and Bain (2008), IBM ViaScribe system was employed when lectures were being given in traditional classroom environment. The lecturer gave a speech on a particular topic and the STR technology simultaneously generated text from the speech which was displayed on the screen for students to read. Students’ questions to the lecturer were repeated by the lecturer to the STR and were transcribed on the screen.

Zschorn et al. (2003) developed Automatic Transcriber of Meetings prototype to use in order to automatically create records and transcripts of a discussion during the meeting. The prototype creates transcripts that include all the attendee, agenda, highlights and utterances information of the meeting.

The STR technology, developed by Augmented Multiparty Interaction with Distance Access, IBM, International Computer Science Institute and SRI International and Karlsruhe University, was employed by Fiscus et al. (2007) at small conference room meetings, interactive lectures in a small meeting room, and coffee breaks from lecture meetings. Conference meetings consisted of primarily goal-oriented, decision-making exercises and varied from moderated meetings to group consensus-building meetings. Conference meetings were highly interactive and multiple participants contributed to the information flow and decision-making. Lecture meetings consisted of educational events where a single lecturer briefed audiences on a particular topic. While the audience occasionally participated in question and answer periods, the lecturer predominately controlled the meeting. Coffee breaks from lecture meetings consisted of excerpts selected from lecture meetings where the participants took a coffee break during the recording.

Fiscus et al. (2007), Li et al. (2011), Wald and Bain (2008), and Zschorn et al. (2003) focused primarily on technological aspects of STR technology. That is, these studies developed STR systems and tried to improve STR accuracy rate, but they didn’t put much emphasis on evaluating the effects of systems on learning achievement and other pedagogical issues such as systems’ practicality or functionality to enhance learning.

Kuo et al. (2011), Kuo et al. (2012), and Shadiev (2011) applied Windows Speech Recognition built in Microsoft Operating System for individual oral presentations and group discussions of native speakers of Chinese students in a synchronous cyber classroom. The effectiveness of applying STR on learning performance was analyzed. Students in the study of Kuo et al. (2011), Kuo et al. (2012), and Shadiev (2011) Kuo et al. (2012) participated in individual oral presentations and group discussions, in which STR technology generated transcripts and it was shown on students’ computer screens. STR-generated transcripts were used by students during and after learning activities.

Potentials and findings of the STR to facilitate learning

This section illustrates related literature review on potentials and findings of the STR to facilitate learning in terms of different users and learning environments. Besides, what considerations associated with STR technology application are reported in this section. Main findings of this section are summarized in Appendix.

Students with cognitive or physical disabilities

Nisbet and Wilson (2002), and Nisbet et al. (2005) evaluated STR application in a classroom by using Pupil Evaluation Questionnaire which was completed by students with special needs, in collaboration with teachers. Students’ responses to the questionnaire were analyzed and results showed that 70% of the students intended to
continue using STR for learning purpose. According to students, STR technology served as an effective tool to write and record their speech. In some cases, the application of STR has enhanced students’ basic reading, spelling and writing skills. For example, the system could read out their STR-text and play recorded audios so that students could compare the read-out of STR-text to playback to identify misrecognitions. Moreover, recordings of the student’s dictations were saved, so that students could correct them later with the help of a teacher.

Leitch (2008), Wald (2008), Wald (2010), and Wald and Bain (2008) conducted a survey on students with hearing impairments. The survey data analysis revealed that STR-generated lecture transcriptions helped students to understand lectures content better. Besides, students believed that transcriptions could improve their learning.

To reveal advantages of STR in learning, Elliot et al. (2002) interviewed students with hearing impairments. High school students claimed that reviewing notes helped them to fill in gaps of their understanding of what transpired during class. College students mentioned that STR notes were useful for test preparation and for traditional academic purposes, e.g., background material for research papers.

Online students

Hwang et al. (2010) and Hwang et al. (2012) carried out an experiment, and its results showed that, in an online synchronous learning environment, students who used transcripts (the experimental group) showed a more moderate improvement in their performance than students who did not use transcripts (the control group) on homework accomplishments. However, once the students in the experimental group familiarized themselves with the STR-generated texts and used them as learning tools, they significantly outperformed the control group students in post-test results. Results of the other experiment, carried out by Kuo et al. (2011) and Kuo et al. (2012) in an online synchronous learning environment showed that students who used transcripts performed significantly better than those who did not use transcripts in writing essays, intermediate tests, and post-test evaluations. Furthermore, experimental students in both studies (Hwang et al., 2012; Kuo et al., 2012) perceived that STR system was easy to use and useful for academic activities in online synchronous cyber classrooms. Yet, students expressed their positive willingness to use STR system for learning in the future. According to interviews with experimental students, STR texts were useful during and after academic activities to understand presented topics, to catch up on missed/misheard parts in a speech, to take notes, and to complete homework. However, it was the STR low recognition accuracy rate when recognizing homophones that students viewed as the one limitation which required attention (i.e., the words with the same pronunciation but different meanings).

Non-native speakers

The participants in the study administered by Ryba et al. (2006) claimed that STR technology has a potential to be an instructional support mechanism, and there were a number of perceived benefits associated with the STR use. Most non-native speaker students, due to their language barrier and mishearing some important parts of the instructor’s speech, admitted that STR-texts were useful during lectures to follow the instructor and to clarify and to understand lecture content.

Experimental results in the study of Coniam (1999) showed that transcripts generated from speeches of second language speakers by using STR were with significantly lower accuracy rate than those generated from speeches of native speakers. These results were consistent in line with native speakers’ scores; that is, the highest accuracy scores were achieved at the lowest level of analysis, the word level, and the lowest scores at the t-unit, or sentence level of analysis. Furthermore, Coniam (1999) concluded that STR technology is still at early stage of development in terms of accuracy and single-speaker dependency.

Results obtained by Shadiev et al. (in press) and Shadiev et al. (2013) revealed that non-native English participants took advantage of nineteen learning strategies to use STR-generated transcripts during and after seminars in English. Transcripts were used to understand seminar’s topics, to answer seminar’s questions, and to complete summary writing tasks. However, participants employed learning strategies differently. That is, some participants used transcripts effectively by studying them thoroughly, and they used most important parts of transcripts to write summaries along with their own elaborated ideas. On the other hand, some participants performed meaningless
learning behaviors as they studied transcripts superficially and employed copy-and-paste method to complete summary writing tasks. As a result, those participants who employed meaningful learning strategies to use STR-generated transcripts received higher scores for their summaries than those who used undesirable learning strategies. Finally, Shadiev et al. (in press) and Shadiev et al. (2013) found that most non-native English participants perceived that available STR-generated transcripts were useful for their learning during and after a seminar. However, low STR accuracy rate was a problem proposed by some participants. This problem caused their negative perceptions and slightly decreased their perceived acceptance to use STR in the future. Those participants admitted that there was not enough time to receive STR technology training. Furthermore, as non-native speakers of English, they may have strong accent in pronouncing some words or stumbled over them, and this caused many errors in STR-generated text when speaking to STR.

By using eye-tracking technique to explore non-native English speaker students’ visual attention to STR-generated text, Shadiev et al. (2014) found that students relied on STR-texts more than on video of the instructor and Power Point slides during lectures in English. Shadiev and his colleagues concluded that STR-texts were useful during the lectures as to aid learning. Students made a greater use of STR-texts to enhance their comprehension of the lectures content. Shadiev et al. (2014) found that all students, no matter what level of their English as a foreign language (EFL) ability, learning style preference and gender are, learned with the aid of STR-texts. However, STR-texts significantly helped to enhance learning performance of participants with low level of EFL ability. Shadiev et al. (2014) argued that participants of low EFL ability took better advantage of STR-texts while being engaged in perceptual processing during listening. For example, some participants admitted that reading STR-texts could help them understand lecture content better. Some participants mentioned that STR-texts could help them to locate new and unfamiliar vocabulary. Results of this study also revealed that participants tended to gaze on all areas of interest, i.e., video of the instructor, Power Point slides and STR-texts during an intermediate-level lecture, but more on STR-texts. Furthermore, results showed that participants tended to gaze mostly on STR-texts during an advanced-level lecture. Shadiev et al. (2014) explained this finding out of difficulty of the lectures; as difficulty of the lecture is higher, participants paid their visual attention to STR-texts more in order to comprehend the lecture content better.

Weggerle et al. (2009) found that introducing STR into the classroom had several positive learning benefits. For example, pronunciation and correct grammar of the lecturer improved substantially and thus, improved students learning. A transcribed text from lectures was voted by students to be very valuable for exam preparation. However, Weggerle et al. (2009) reported that employing STR technology in their study rarely offered a recognition rate of more than 80 percent, and the delay involved in real time transcription was disturbing. According to the literature on STR (Hwang et al., 2010; Kheir & Way, 2006; Wald, 2010), text generated under such circumstances becomes unhelpful and meaningless for students’ learning.

**Students in traditional learning environment**

Ranchal et al. (2013) concluded that during a science course in traditional classroom, students could benefit from having both, real-time lecture transcriptions and post lecture transcriptions. When lecture transcripts were available, students were able to pay more attention to the instructor instead of focusing on recording complete class notes, and with the lecture transcripts, they could review the lecture material for several times. Besides, students were able to take notes, make comments and remarks, and look for specific text by searching keywords and time periods. However, Ranchal et al. (2013) found that students who had access to post lecture transcriptions received higher scores on the quiz than those who received real-time transcriptions only. Moreover, overall class grades of students who received post lecture transcriptions were higher.

Results of the class survey in the study of Ryba et al. (2006) revealed that more than 30% of students used STR-texts to learn information systems in a traditional classroom. Ryba et al. (2006) further found that more than 40% of students tend to use STR-texts. In the survey, students mentioned that STR-text helped them to understand the lecture, confirming what was missed in the lecture, and to take notes. However, most students claimed that the accuracy rate of STR technology was not precise enough, and text generated with many errors could distract their attention from the lecture.

Goddard et al. (2007) surveyed their participating teachers and primary general education students about benefits of Read & Write GOLD system. From the survey, it was found that the system benefits students to write, to edit, and to rewrite. According to the teachers, students’ writing improved after they started using the system. Students heard and
recognized obvious errors that they, at first, did not believe they had made as the system read exactly what the students had written. Editing was not a struggle as the software was reading students’ work to them. Teachers reported that spending more time with writing, editing, and rewriting improved the final product. Furthermore, teachers continuously reported that all students were engaged in using the software throughout the year, not just short-term interest.

**The STR technology applications in collaborative academic activities**

Kuo et al. (2012) found that STR technology is a potential tool to facilitate collaborative learning activities, such as oral presentations and group discussions, and it can also improve their overall learning performance. Experimental results in the study of Kuo et al. (2012) and Shadiev (2011) revealed that students who used STR-generated texts (the experimental group) performed far better than those who did not (the control group) in writing essays, intermediate tests and post-tests. Furthermore, according to results, most students perceived that STR was a useful aid when prepare for oral presentations and essays writing. However, there was a problem that it was difficult to attain a high recognition accuracy rate of STR during group discussion. Therefore, students who got transcripts with low accuracy rate and experienced delay in STR-text generation did not perceive STR as an easy tool to use, and found it not so useful for group discussions. One reason of having a low accuracy rate was due to a speed of students’ speech. When a student spoke too slowly, the STR application recognized one spoken word as two. Conversely, when the student spoke too quickly, the STR application recognized two spoken words as one. Furthermore, it was not easy to attain a fluent speech (i.e., when the speech has to be delivered moderately fluent and accurate) during group discussion so that the STR generated texts with low accuracy rate. In addition, students mentioned that their speech became more spontaneous during group discussion which also resulted in low accuracy of transcriptions content. Due to these issues, students couldn’t make argumentative discourse with the goal to acquire knowledge but were engaged in idea exchange only.

Literature review shows that participants in most studies on STR, no matter what category of users they belong to and no matter what learning environment they learn in, had positive perceptions toward usefulness of STR transcripts for learning. However, Mayer (2008) argued that the same information presented in both auditory and written format makes it redundant and gives rise to a split-attention effect and cognitive load (modality principle). However, the participants still relied on transcripts in written format because of their learning needs, physical/ cognitive abilities, or specific learning environment (Elliot et al., 2002; Hwang et al., 2012; Kuo et al., 2012; Leitch, 2008; Nisbet & Wilson, 2002; Nisbet et al., 2005; Ryba et al., 2006; Shadiev et al., 2013; Shadiev et al., 2014; Wald, 2010; Wald & Bain, 2008). According to Kirsh (2010), and Rogers, Sharp, and Preece (2011), external representations, such as STR-generated texts, greatly extend and support students’ ability to carry out cognitive activities (e.g., inference, problem-solving and understanding). One benefit that transcripts offer is on memory. Firstly, transcripts reduce memory workload by providing external tokens for the information that must otherwise be kept in mind. Secondly, transcripts serve as visual retrieval cues for long term memory, evoking relevant information that might not otherwise be retrieved. Finally, transcripts are more “enduring” (visual) text-based content, which goes along with the more “temporary” (oral) speech-based presentation. According to Dual Processing theory (Moreno & Mayer, 2002), redundant information presented in two modes (i.e., visual and oral), and processed aurally and visually can support the recognition and learning of that information. Thus, in the finding of Moreno and Mayer (2002), participants used strategies such as scanning transcriptions when they missed or misheard some parts of a speech. In this way, STR technology can provide much more essential support for students to process aural text with the help of simultaneously displayed transcriptions (Jones & Plass 2002; Ryba et al., 2006).

**The STR considerations**

Three main issues with respect to the STR technology were pinpointed by teachers and students in the reviewed literature. First issue was reported in Hwang et al. (2012), Kuo et al. (2012), Shadiev et al. (2013), and other related studies and it relates to the usage of the STR technology. It was found that students who did not use the STR technology or used it irregularly perceived STR not a useful aid for learning. The second issue associates with STR process accuracy rate. Most studies report that although STR technology is useful for learning, a greater accuracy in the system’s recognition of speech is required. According to Alapetite et al. (2009), Fichten et al. (2000), Jones (2005), Kanevsky et al. (2006), Kheir and Way (2006), Konur (2007), Petta and Woloshyn (2001), and Wald (2010), texts generated with low accuracy recognition rate contain many errors which are incomprehensible and meaningless for learning. Finally, it’s the issue that relates to learning behaviors in using STR texts. Shadiev and his colleagues
(2013) noticed that participants in their study performed slack learning behaviors, such as studying transcripts superficially and employing copy-and-paste method to complete summary writing tasks. Performing such learning behaviors, students did not learn much, and as a result, they were scored low on examinations.

Suggestions and implications

To begin with, the literature review suggests that educators and researchers design technology-based teaching and learning activities in a way that encourages users (i.e., instructors and students) to use STR more regularly. Such approach will enable users to identify strengths and limitations of the STR, and then to fully utilize STR for their teaching and learning. For example, Hwang et al. (2012), Kuo et al. (2012), and Shadiev et al. (2013) encouraged and motivated their participants by training them how to use STR technology first and then to use it to complete homework. With such kind of learning activity design, students could identify what advantages and disadvantages of the STR are through real experience with STR technology.

According to Hwang et al. (2012), Jones (2005), Kuo et al. (2012), and Nisbet, Wilson, and Balfour (2008), in order to achieve good detection accuracy rate, the STR application training should last at least one week. Hwang et al. (2012) and Kuo et al. (2012) argued that by using training scripts with content related to the learning material, STR technology can “learn” domain-specific terminologies during the training period and then it can recognize them when learning activities are ongoing. To increase STR process accuracy rate during the training period and academic activities, Nisbet et al. (2008) suggested that we use STR dictionary and correction tool. For example, according to Ranchal et al. (2013), a user can add words that are frequently detected to the dictionary so that STR recognizes those words easier. Besides, Ranchal et al. (2013) claimed that a user can simultaneously correct errors in transcript while speaking to STR by using the STR correction tool. Furthermore, recognition errors can be corrected after the lecture. In this case, the instructor or teaching assistant listens to the lecture audio recording and corrects misrecognized words, inserts missed words, or deletes superfluous wording (Ranchal et al., 2013). If transcripts were generated with high error rates, students in the class can be involved in this work collectively by using an online correction tool and sharing the workload among several people. The STR correction tool can also help to train STR against a word that consistently misrecognized; for that, a user has to record a pronunciation of how he/she says that word. Hwang et al. (2012) and Kuo et al. (2012) also suggested that it is feasible to apply a set of strategies during the training on STR technology. Such strategies involve sharing issues related to the STR process with peers, finding possible solutions together, preparing a script with main points of a speech and making rehearsal with a script and STR technology beforehand. According to Colwell et al. (2005), Hwang et al. (2012), Kheir and Way (2006), Kuo et al. (2012), and Wald (2010), only STR-generated text with reasonable accuracy rate of more than 85 percent is useful and meaningful for students. Kheir and Way (2006) reported that, in their study, the accuracy rate of STR improved from 75 percent, when STR was not trained, to 88 percent after minimal training on STR, to 90 percent after moderate training, and to 91 percent after its dictionary was customized with a domain-specific terminology. Furthermore, Kuo et al. (2012) and Ranchal et al. (2013) suggested that speakers try to adapt to the STR recognition capacity by speaking with moderate speed and volume, less spontaneity, and better fluency. Microphone should also be positioned correctly to avoid “breathiness.” Nisbet et al. (2008) suggested a speaker speak clearly to STR and avoid non-lexical utterances (e.g., “huh,” “uh,” or “erm”). Only the speaker’s voice should be reliably recorded; if responding to students’ questions, the instructor should repeat questions and then respond (Ranchal et al., 2013). To increase its accuracy rate during discussion, Kuo and his colleagues advised that speakers make speaking sentences shorter and at a moderate pace of one sentence after another, and to locate and correct errors in the transcript simultaneously while speaking to STR. Ranchal et al. (2013) recommended speakers to take breaks periodically if lectures are long to check the reliability of the STR system. Based on these findings, it is suggested that offering users a set of guidelines on how to train and speak to STR more efficiently can achieve better STR accuracy rate and make transcripts more useful and meaningful for learning.

To avoid students’ slack learning behaviors, it is suggested that, besides providing learning material, participants need to be instructed about how to use effective learning strategies to use STR-texts. Learning strategies to use STR-generated texts were proposed in Nisbet & Wilson, 2002; Nisbet et al., 2005; Ryba et al. (2006), Shadiev et al. (2013), and Shadiev et al. (2014). These strategies can facilitate participants during and after an academic activity to understand content of a presented topic better, to answer questions, and to complete summary writing tasks. Shadiev
et al. (2013) and Shadiev et al. (2014) suggested some more advanced strategies to the ones reviewed in related literature. Two of them are 1) to use a transcript to ask questions, to give comments or to have discussion with others and 2) to compare a transcript with a student’s summary in order to confirm that a summary includes all main points of a speech.

Finally, it is suggested that STR technology can be applied in a learning environment not only with a single speaker but with multi speakers as well. In this case, individual learning, such as using lecture transcript to involve in a speech, taking notes and completing homework, will be enhanced. That is, after individual learning, students can share and discuss their opinions about the topic, correct each other’s misconceptions, and enhance their own understanding of a topic by using STR technology. However, some issues need to be considered with respect to STR technical and pedagogical process in such learning scenario. For example, one is how to make STR correctly recognize speech input made by multiple speakers with different speech characteristics (e.g., articulation, pronunciation, speech rate) and then distinguish that input in a STR-generated transcript by each speaker and with orderly timeline when it was spoken (Fiscus et al., 2007; Li et al., 2011; Wald & Bain, 2008; Zschorn et al., 2003). Another issue is how to design collaborative learning activities that facilitate students to fully utilize STR-generated texts for learning (Hwang et al., 2012; Kuo et al., 2012).

Conclusions

The following conclusions can be drawn from literature review. First, it is fairly clear that STR technology was applied to aid learning in different ways based on the progress of STR technological development. That is, the earlier studies employed this technology only to assist learning of particular groups of users, such as students with cognitive and physical disabilities, online or foreign students, due to low accuracy rate and a delay in STR process. However, afterwards, research addressing abovementioned technological limitations emerged; as a result, STR technology improved and became more reliable. Then, STR was employed to aid learning of students in a traditional learning environment during and after individual and collaborative learning activities. Second, the literature considered STR technology beneficial to extend learning during and after learning activities. There is widespread consensus in the literature about the number of distinct advantages of STR-texts, such as enabling students to better understand content of academic activities, to confirm missed parts of a speech, to take notes, to complete homework, and to prepare for exams. However, some arguments over the STR technology considerations that limit educational value of the technology still exist. The literature review showed how those considerations can be addressed by employing various approaches to increase the effectiveness of STR application on learning.

Given what was found in the literature, the following are important issues to address in future STR related studies. First, researchers need to begin theorizing the cognitive processes that occur through learning with STR technology. Besides, STR technology needs to be employed based on relevant pedagogical principles for them to be more effective. Second, there is a need to use well-established and reliable outcome measures in future STR studies. For example, the measures used to demonstrate the effects of STR applications should be given careful consideration based on both objective and subjective evidence. More research needs to be conducted in more dynamic and communicative educational settings, such as collaborative teaching and learning with multiple participants speaking to the STR system simultaneously. Besides, whether there will be different learning effects when the STR is applied to learning environments with students of different cultural backgrounds or language families should be investigated. It is possible that there might be impacts on the STR accuracy rate and learning when different cultural backgrounds or language families are concerned. For example, in general, students from oriental cultural background are less active in terms of learning interaction which may influence learning dynamics and outcomes during group discussion. Finally, research should focus on issues that go beyond applications of STR technology. For example, STR technology can be considered from the angle of ergonomics, i.e., to concern the design and arrangement of the technology to make users interact with it more efficiently. Besides, in the future, STR technology can be extended by combining it with other technology, e.g., automatic translation, to simultaneously generate text from a speech and translate it into many languages. Such approach will enable teachers and students to have instant audio-lingual interpretations using their own native languages.
Acknowledgments

This research is partially supported by the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301, NSC 102-3113-P-006-019-, MOST 103-2511-S-006-007-MY3, and MOST 103-2511-S-006-002-MY3.

References


77


References, which content was reviewed for the analysis and deriving findings, are marked with an asterisk.
### Appendix

#### Research findings on applications of STR to enhance learning

<table>
<thead>
<tr>
<th>Reference</th>
<th>Research focus</th>
<th>Target group</th>
<th>STR methodology</th>
<th>STR system</th>
<th>General findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Students with cognitive or physical disabilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nisbet &amp; Wilson (2002)</td>
<td>To investigate best practices of STR applications in schools.</td>
<td>Secondary school students with reading, writing and spelling difficulties.</td>
<td>Students individually used STR system during class; STR-text was simultaneously displayed to students.</td>
<td>Dragon Naturally Speaking / IBM ViaVoice</td>
<td>- Most students intended to continue using STR for learning purpose; - STR was found as an effective tool to write and record; - In some cases, STR has enhanced basic reading, spelling and writing skills.</td>
</tr>
<tr>
<td>Nisbet et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leitch (2008)</td>
<td>To understand whether applications of STR can assist in creating a positive and beneficial learning environment for students.</td>
<td>High school students with hearing impairments.</td>
<td>The instructor applied STR during lectures; Lecture transcription was simultaneously displayed to students on a whiteboard/computer screens.</td>
<td>IBM ViaVoice</td>
<td>- STR-texts helped students to understand lectures content better; - Students believed that STR-texts could improve their learning.</td>
</tr>
<tr>
<td>Wald (2010)</td>
<td>To understand how STR applications may contribute to an improved learning environment for students.</td>
<td>University students with hearing impairments.</td>
<td>The instructor applied STR during lectures; Lecture transcription was simultaneously displayed to students on a whiteboard/computer screens.</td>
<td>IBM ViaVoice</td>
<td>- STR-texts helped students to understand lectures content better; - Students believed that STR-texts could improve their learning.</td>
</tr>
<tr>
<td>Wald &amp; Bain (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elliott et al. (2002)</td>
<td>Students’ learning strategies to study with STR notes were explored.</td>
<td>High school and college students with hearing impairments.</td>
<td>The instructor used STR to pre-generate lecture notes; Lecture notes were delivered to students.</td>
<td>C-Print</td>
<td>- STR-generated notes helped high school students to fill in understanding gaps; - STR notes were useful for college students to prepare for the test and to write research papers.</td>
</tr>
<tr>
<td>2. Online students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hwang et al. (2010)</td>
<td>The effectiveness of STR applications on students</td>
<td>Open university</td>
<td>The instructor</td>
<td>Windows Speech</td>
<td>- Experimental students perceived that STR</td>
</tr>
<tr>
<td>Research</td>
<td>Methodology</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hwang et al. (2012)</td>
<td>Students applied STR during lectures; Lecture transcription was simultaneously displayed to students online; The instructor provided students with edited transcriptions after lecture.</td>
<td>Recognition system was easy to use and useful for one-way lectures and individual learning; Most experimental students expressed that they were highly motivated to use STR as a learning tool in the future; Experimental students performed moderately better compared to control students in homework accomplishments; Experimental students significantly outperformed control students in post-test results.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shadiev (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuo et al. (2011)</td>
<td>Students who used STR-texts (experimental) outperformed students who did not use STR-texts (control) on essays writing, intermediate test and post-test; Most experimental students perceived that STR was useful for individual presentations and for essays writing; Experimental students were willing to use STR system for learning in the future; Experimental students, who obtained transcripts with low accuracy rate and experienced delay in STR-text generation, perceived STR system wasn’t easy to use and useful for group discussions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuo et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shadiev (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Non-native speakers</td>
<td>University students (native and non-native speakers) applied STR during lectures; STR-text was displayed on a whiteboard;</td>
<td>IBM ViaVoice - STR technology has a potential to be an instructional support mechanism; Most non-native speaker students admitted that STR-texts were useful to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ryba et al. (2006)</td>
<td>The instructor applied STR during lectures;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coniam (1999)</td>
<td>Potentials of STR applications to enhance students’ learning English as a second language was explored.</td>
<td>Second language learners (L2)</td>
<td>Dragon Naturally Speaking</td>
<td>- L2 learners generated texts from their voices by using STR system; - STR-texts of L2 learners was analyzed and compared with STR-texts of native speakers. - STR-texts of L2 learners were less accurate compared to those of native speakers in each category of analysis; - The highest accuracy scores were achieved at the lowest level of analysis, the word level, and the lowest scores at the t-unit, or sentence level of analysis.</td>
<td></td>
</tr>
<tr>
<td>Shadiev et al. (in press)</td>
<td>Students’ perceptions toward STR applications, the difference between using one-week summaries versus immediate summaries, and learning behaviors to use STR-text were studied.</td>
<td>Graduate students (non-native speakers)</td>
<td>Windows Speech Recognition in the Microsoft Operating System</td>
<td>- Students applied STR during seminar; - Transcripts generated during seminar were simultaneously displayed to students online; - A speaker provided students with edited transcriptions after seminar. - Nineteen learning strategies to use STR-texts were revealed; - Participants employed learning strategies differently; - Participants scored differently in their summary writing assignments; - Most participants perceived that STR-texts were useful for learning; - Low accuracy rate was a problem proposed by some participants.</td>
<td></td>
</tr>
<tr>
<td>Shadiev et al. (2013)</td>
<td>Visual attention on STR-text, how differently effective STR-texts can be to influence learning achievement, and students’ perceptions regarding usefulness of STR-texts for learning were investigated. Furthermore, visual attention and learning behaviors to different participants’ characteristics (i.e., learning ability, learning style preferences, and gender) to use STR-text were compared.</td>
<td>Graduate and undergraduate students (non-native speakers)</td>
<td>Windows Speech Recognition in the Microsoft Operating System</td>
<td>- STR-texts were displayed to students on computer screens during two lectures on intermediate and advanced levels. - Participants relied on STR-texts more than on video of the instructor and Power Point slides; - Participants made a greater use of STR-texts to enhance their comprehension of the lectures content; - Participants, no matter what levels of their EFL ability, learning style preference and gender are, learned with the aid of STR-texts; - STR-texts significantly helped to enhance learning performance of low ability participants.</td>
<td></td>
</tr>
<tr>
<td>Weggerle</td>
<td>To enhance the learning University</td>
<td>Dragon</td>
<td>The</td>
<td>- With the help of STR</td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Study Description</td>
<td>Students (Language)</td>
<td>Instructor Used STR System During Lectures</td>
<td>Naturally Speaking System, the Lecturer Pronunciation and Correct Grammar Improved Substantially and Thus, Improved Students Learning;</td>
<td>STR System, the Lecturer Pronunciation and Correct Grammar Improved Substantially and Thus, Improved Students Learning;</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>et al. (2009)</td>
<td>Efficiency by applying STR system during lectures.</td>
<td>students (foreign language students and non-attendants)</td>
<td>instructor used STR system to generated texts from voice input; STR-texts were edited and provided to students as lecture notes.</td>
<td>Students perceived that STR-texts are very valuable tool for exam preparation; STR recognition rate rarely was obtained more than 80 percent, and the delay in real time transcription was disturbing.</td>
<td>Students perceived that STR-texts are very valuable tool for exam preparation; STR recognition rate rarely was obtained more than 80 percent, and the delay in real time transcription was disturbing.</td>
</tr>
<tr>
<td>4. Students in traditional learning environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranchal et al. (2013)</td>
<td>The effectiveness of real-time captioning and post-lecture transcription on learning were evaluated.</td>
<td>University students</td>
<td>- The instructor applied STR system during lectures; - Real-time lecture transcriptions streamed on computer screen; - Edited STR-texts were provided to students after class.</td>
<td>Students benefited from both, real-time lecture transcriptions and post lecture transcriptions; Real-time lecture transcriptions helped students to pay more attention to the instructor, to take notes, make comments, remarks and dynamically search for specific lecture keywords and time periods; Students who had access to post lecture transcriptions received higher scores on the quiz compared to students who received real-time transcriptions; Overall class grades of students who received post lecture transcriptions were higher.</td>
<td>Students benefited from both, real-time lecture transcriptions and post lecture transcriptions; Real-time lecture transcriptions helped students to pay more attention to the instructor, to take notes, make comments, remarks and dynamically search for specific lecture keywords and time periods; Students who had access to post lecture transcriptions received higher scores on the quiz compared to students who received real-time transcriptions; Overall class grades of students who received post lecture transcriptions were higher.</td>
</tr>
<tr>
<td>Ryba et al. (2006)</td>
<td>Students’ usage of STR-texts and perceptions toward usefulness of STR-texts for learning were explored. Moreover, the advantages/disadvantages of using STR-texts for learning were investigated.</td>
<td>University students</td>
<td>- The instructor used STR system during lectures; - STR-texts were simultaneously displayed on a whiteboard; - Students obtained edited IBM ViaVoice and Hosted Transcription Service</td>
<td>More than 30% of students used STR-texts for learning; - STR-texts were useful to understand lectures, to confirm what was missed in lectures, and to take notes; - Most students complained that the accuracy rate of STR technology was too low.</td>
<td>More than 30% of students used STR-texts for learning; - STR-texts were useful to understand lectures, to confirm what was missed in lectures, and to take notes; - Most students complained that the accuracy rate of STR technology was too low.</td>
</tr>
<tr>
<td>Study</td>
<td>Description</td>
<td>Participants/Environment</td>
<td>Findings/Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goddard et al. (2007)</td>
<td>How teachers and students use STR system in classroom to support students’ learning needs was investigated.</td>
<td>Primary school students</td>
<td>- STR-texts after lectures. &lt;br&gt; - Students trained STR system to their voices;  &lt;br&gt; - Students spoke to STR system and texts were generated from their voices;  &lt;br&gt; - Students’ speeches were audio recorded;  &lt;br&gt; - System read back generated texts and played back students audio recordings.</td>
<td>- Writing, editing, and rewriting were classroom benefits of the system for students’ compositions;  &lt;br&gt; - Students’ writing improved after they started using the system;  &lt;br&gt; - The system helped students to hear and recognize errors that they made;  &lt;br&gt; - Spending more time with writing, editing and rewriting improved the final product.</td>
<td></td>
</tr>
<tr>
<td>Wald and Bain (2008)</td>
<td>To understand how STR system may contribute to an improved learning environment.</td>
<td>University students</td>
<td>- The instructor used STR system during lectures; &lt;br&gt; - STR-texts were simultaneously displayed on a whiteboard; &lt;br&gt; - Students asked questions which were repeated by the lecturer to STR system so questions also appeared transcribed on a whiteboard.</td>
<td>Focus on technological aspects of STR technology: STR system was developed and researchers attempted to improve its accuracy rate. Effects of the system on learning achievement and system’s practicality or functionality in pedagogical aspect were not evaluated.</td>
<td></td>
</tr>
<tr>
<td>Zschorn et al. (2003)</td>
<td>To develop and evaluate STR system that produces text and audio records of a discussion during meetings.</td>
<td>General group of users</td>
<td>- As the meeting participants speak, STR system generates texts from voice inputs and segments speeches into Automatic Transcriber of Meetings prototype</td>
<td>Focus on technological aspects of STR technology: STR system was developed and researchers attempted to improve its accuracy rate. Effects of the system on learning achievement and system’s practicality or...</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Purpose</td>
<td>Participants</td>
<td>Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Fiscus et al. (2007) | To design and evaluate the Rich Transcription Meeting Recognition.     | General group of users                                                        | - As meeting participants speak STR system transcribes voice inputs into texts.  
Focus on technological aspects of STR technology: STR system was developed and researchers attempted to improve its accuracy rate. Effects of the system on learning achievement and system’s practicality or functionality in pedagogical aspect were not evaluated. |
| Kuo et al. (2011)   | The effectiveness of applying STR during collaborative learning activities on learning performance was analyzed. | Open university students                                                      | - Students used STR system for oral presentations and group discussions;  
- Speakers took turns to speak;  
- STR system generated texts from voice inputs and displayed them simultaneously on computer screens;  
- STR-texts were available to students after learning activities.  
- Applications of STR could facilitate collaborative learning activities as to improve students overall learning performance;  
- Students who used STR-texts (experimental) outperformed students who did not use STR-texts (control) in two sessions of writing essays, intermediate test and post-test;  
- Most experimental students perceived that STR system was useful for individual presentations and for essays writing;  
- Experimental students expressed their willingness to use STR system for learning in the future;  
- Experimental students who obtained transcripts with low accuracy rate and experienced delay in STR-text generation did not perceive STR system as easy to use and useful for group discussions. |
The Influence of Young Children’s Use of Technology on Their Learning: A Review

Ching-Ting Hsin¹, Ming-Chaun Li²* and Chin-Chung Tsai³

¹Department of Early Childhood Education, National Taitung University, Taiwan // ²Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan // ³Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // cthsin@nttu.edu.tw // mli1tw@gmail.com // cctsai@mail.ntust.edu.tw

*Corresponding author

ABSTRACT

This study aimed to conduct a systematic literature review on empirical studies of how technologies influence young children’s learning. Eighty-seven articles published between 2003 and 2013 were identified through the Web of Science database. We employed content analysis to identify the research trends of this topic. “Technology evaluation,” “adults’ roles,” and “teaching approaches” are three emerging research themes during 2008-2013. About one-third of the studies involved children who were from immigrant or low socioeconomic status families, or had special needs. The majority of the reviewed studies revealed that the technologies had positive effects on children’s performance across developmental domains. Particularly, in social domain, most studies showed that technologies enhanced children’s collaboration and interaction with others and their development of multiculturalism. We also propose a typology for conceptualizing the complexity of the relationships between technology use and children’s learning. We argue that children’s learning with technology is conditioned by several factors categorized into children, adults, and technology aspects. Moreover, a trend of examining children’s development of digital literacy emerged, involving investigation of the skills needed for and perceptions of technology use. Lastly, while most studies viewed children as consumers of technology, their role as creators has been understudied and deserves more research attention.

Keywords

Early childhood education, Young children, Educational technology, Technology-assisted learning

Introduction

Contemporary young children are part of the generation of digital natives (Fleer, 2011; Prensky, 2001a, 2001b). Young children in this study refer to children aging from 0 to 8 years (Bredekamp & Copple, 1997). They live in a world enveloped by technologies and use technologies in their daily life (Hague & Payton, 2010; Plowman, Stevenson, Stephen, & McPake, 2012). Many countries recognize the increasing role of technology in children’s lives. They emphasize the development of technology-integrated curricula that are developmentally appropriate for young children and that help to bridge young children’s digital experiences at home and in school (Mawson, 2003; McKenney & Voogt, 2009; Plowman, Stevenson, McPake, Stephen, & Adey, 2011).

Because of the rapid development of technologies, they have changed children’s lives and ways of learning, particularly in the past ten years. Researchers have urged a rethinking of the roles of technology in young children’s development and consequently the development of learning theories and curricula that meet the needs of contemporary children (Fleer, 2011; Yelland, 2011). Although many researchers and educators have advocated for the importance of young children’s learning with technology and devoted themselves to investigating and implementing technology-related practices, the influence of young children’s use of technologies on their development is still controversial. Some researchers believe that the use of technologies may impede these children’s social, emotional, physical, and cognitive development (e.g., Armstrong & Casement, 2000; Cordes & Miller, 2000), while others support the use of technologies in improving young children’s development in the aforementioned domains (e.g., Clements & Sarama, 2003; Plowman & McPake, 2013; Plowman & Stephen, 2003; Yelland, 2011).

Such discussion surrounds one question about which early childhood educators have been concerned: are the technology-related practices developmentally appropriate for young children (Radich, 2013)? In terms of developmentally appropriate practices, knowledgeable adults play important roles in scaffolding young children’s learning within the zone of proximal development (Bredekamp & Copple, 1997; Vygotsky, 1978). Researchers are thus concerned with the learning effect on children between adult-facilitated and technology-assisted learning. For example, de Jong and Bus (2004) conducted a study to compare children’s learning outcomes after they listened to
adults’ storybook reading and read e-books on their own. Another example of developmental appropriate practice is that children learn abstract concepts through manipulating concrete objects (Dunn, 2001; Hsin, 2012). Researchers have thus debated the effects of manipulating physical materials and virtual materials on children’s learning of science or mathematic concepts (Clements & Sarama, 2003; Zacharia, Loizou, & Papaevripidou, 2012). Moreover, promoting the development of social skills is considered one of the important developmentally appropriate practices for young children (Bredenkamp & Copple, 1997). Some researchers have argued that technology may impede young children’s social skills because children develop these skills through in-person interaction, and their use of various technologies keeps them from such interaction (Armstrong & Casement, 2000). In contrast, some researchers have indicated that technology in fact promotes children’s social development in various ways (Infante et al., 2010). Researchers on both sides argue that the practices they advocate are developmentally appropriate. However, such binary discussion can lead researchers and educators to overlook the complex relationships between children’s use of technology and their learning. We therefore aim to provide a typology for effectively conceptualizing the interplay among critical factors that influence children’s learning with technology.

Another concern which motivated our examination of the relationships between technology use and children’s learning is a lack of a complete, in-depth picture of the past ten years that shows (a) how technologies play a role in children’s learning across the aforementioned four developmental domains, and (b) what research themes and methods researchers have focused on. Although researchers have debated on and raised the importance of this topic for the past ten years, there has been little attention given to a systematic literature review of the empirical studies that have been conducted to understand young children’s learning with technology in different developmental aspects. Also, an overview of the research purposes and methods of these empirical studies is needed. A more complete picture of this topic would encourage researchers to fill the research gaps and to address issues that have not been fully elaborated or supported with evidence. It would also consequently support the development of technology-integrated curricula.

Researchers have paid attention to not only how technologies affect young children’s learning across domains, but also how young children learn to use a variety of technologies, that is, the development of digital literacy. In comparison with the traditional view of literacy, such new forms of literacy emphasize children’s abilities to comprehend and create multimodal digital texts in order to communicate with texts or others (Bawden, 2008; Lankshear & Knobel, 2008). There has been an emergent research trend in digital literacy. However, how preschoolers and kindergarteners develop their digital literacy and how they enact their roles as creators rather than consumers (Taylor, 1980) of technologies have, as yet, been understudied.

In response to the aforementioned needs, we have conducted a systematic literature review and initiated an evidence-based discussion on how technologies influence young children’s learning. We asked the following research questions:

- How do technologies influence young children’s learning across different developmental domains?
- What are the purposes and methods focused on by researchers when conducting studies of this topic?
- What are the key factors that influence children’s learning with technology?

**Method**

**Paper selection**

In this review, the Web of Science database was used to search for articles regarding young children’s learning with technology published from 2003 to 2013. The search of the literature was carried out in May 2013. The reviewed articles were identified through the following procedures. First, a set of technology related keywords was used in combination with the other set of keywords regarding target age groups by employing the Boolean operator “AND.” The technology related keywords were technology/technologies, computer/computers, media, medium, multimedia, digital literacy/literacies, and multimodal/multimodality. The keywords related to target age group were young child/children, preschooler/preschoolers, kindergarten/kindergartner, early child/childhood, and early year/years. The Boolean operator “OR” was utilized to combine the keywords within the same set. In addition, the truncation search technique was adopted to cover the variations of keywords. For example, technology was used to search for literature that included the word technology or technologies in the target search fields. Because the emphasis of this review is on children’s learning, the results of the keywords search were limited to the “Education Educational
Research” category in the database. Moreover, only peer-reviewed journal articles were selected to ensure the quality of the studies reviewed. These articles are all indexed in Social Science Citation Index (SSCI). Lastly, targeted articles had to be written in English due to a lack of comprehension of other languages. The database search resulted in 273 articles for further selection. It should be noted that the articles selected for the current review are not meant to be inclusive, but are used to explore the emergent issues of this line of research during the past decade. It might not be appropriate to generalize the results of this study to the research field of technology-supported education in early childhood.

Following the database search, the researchers read through the titles and abstracts of the articles to select target papers that met the following criteria: (a) they should be empirical studies that investigated the effectiveness of a technology on children’s learning or surveyed the relationships between technology use and children’s learning, (b) at least one technology is adopted or reported for children’s learning, (c) the age of the participating children is under 8 years old, and (d) the full text of the article should be available either in paper or electronic format. If sufficient information for selecting the articles was not provided in the abstracts, the researchers then went through the major parts of the articles (e.g., methodology and results) to make the judgments. If a study investigated the influence of technology on learning across different age groups, only the instructions and learning results in relation to the children under 8 years old were reviewed. However, the article was excluded if the target children were not the main focus of the study among different age groups. Furthermore, if the main focus of the research was on technology development or instructional design and no essential outcome data were provided, the article was excluded as well. After this selection process, 87 articles were identified for this review. A complete list of the reviewed articles can be downloaded from http://goo.gl/bT7DsK.

**Paper analysis**

The identified articles were analyzed using a content analysis technique. NVivo 10 analysis software was utilized to facilitate the development of the coding scheme. First, 18 articles were randomly selected and read by the authors to develop an initial coding scheme in accordance with the research questions. Next, another 54 articles were randomly selected and assigned to the first and second authors. Each author coded 27 articles independently. When there was a need to add, remove, merge, or modify a code, the authors discussed the code until a consensus was reached. After both authors completed the previous coding process, the remaining 15 articles were coded by both authors independently. The inter-coder agreement was calculated for the coding results of these 15 articles. Agreement reached 87%.

Several rules were applied to the coding task. When there was more than one study reported in the article, only the study/studies that met the review criteria would be coded (e.g., Manches, O’Malley, & Benford, 2010). If all the studies met the review criteria, they were coded separately as individual articles (e.g., Magnan & Ecalle, 2006). If more than one study was conducted in an article but the results were integrated among studies, those studies were coded as one article (e.g., McPake, Plowman, & Stephen, 2013). As a result, 94 studies from 87 articles were analyzed. Participating children in these studies aged from 0 to 8 years. Children’s use of technology refers to their engagement in activities involving various types of technology and concomitant development in various domains. The types of technology identified from these studies include information and communication technology (ICT) (e.g., computer programs and Internet), multimedia (e.g., E-books and TV) and digital devices (e.g., interactive whiteboard and robot).

**Coding scheme**

The coding structure developed through the content analysis consists of five categories, with each including several sub-categories.

**Research purposes**

Six research purposes were identified. If a study involved more than one research purpose, only the major one was coded. These research purposes are described as follows.
Technology evaluation: Investigate the influence of technologies on children’s learning.

Technology vs. tradition: Compare the influence of technology-assisted learning and traditional ways of learning (without technology).

Teaching approaches: Compare the influence of different teaching and learning approaches or models embedded in the design of a technology. For example, a study compared the effects of systematic and unsystematic phonics training on language learning (De Graaff, Bosman, Hasselman, & Verhoeven, 2009).

Mechanism design: Compare the influence of different mechanism designs adopted in the same technology. For example, a study revealed how different interfaces (i.e., individual and paired control) led to different behavior of children (Druin et al., 2003).

Adults’ roles: Investigate adults’ roles in children’s technology use, including adults’ facilitation of children’s engagement in using technology, adaptation of their teaching to respond to children’s learning with technology, and perceptions of children’s technology use.

Consumers vs. creators: Explore the ways in which children learn content delivered by technologies (consumers) or create digital texts and write program to solve problems (creators).

Research design

Five major research designs were identified including: experimental design (with random assignment), quasi-experimental design (without random assignment), one group pre/post-test or post-test only design, case study, and survey. Any research designs other than these five methods were categorized as “Others.” If a study adopted more than one research design, only the major one was coded.

Research participants

Children were grouped into those with a general background and those with diverse backgrounds. Children were categorized as diverse if they were from immigrant or low socioeconomic status families, or had special needs (e.g., reading difficulties, dyslexia, developmental delay, Autism, Down syndrome, or pre-, peri- and post-natal complaints).

Developmental domains

Four developmental domains were used to classify the learning results: the cognitive, social, emotional, and physical domains. Under the cognitive domain, the results were further categorized into different learning areas including language and literacy, digital literacy, math, science, cognitive abilities, and others. When there were multiple results reported in a study, each result was coded into a developmental domain. If the result was related to the cognitive domain, the specific area under the cognitive domain was coded.

Learning results

Each learning result was also categorized into positive, negative, no difference, or depends. When a result was coded as depends, it indicates that the effectiveness of the technology for children’s learning was conditional by various factors (e.g., children’s age, adults’ roles, or mechanism designs of technology). The conditional factors were also coded for further analysis and discussion.

Results and discussion

The analysis results are presented in the following subsections. First, an overview of the descriptive statistics is provided. Second, cross-analysis results between reviewed categories, including research purpose and research design as well as developmental domains and learning effectiveness are reported. Following the quantitative analysis results, the in-depth analyses and discussion are organized into four parts: a typology for conceptualizing young
children’s learning with technology, an emergent discussion of digital literacy, the supports of technology for children’s social development, and children’s roles as creators of technologies.

Overview of the reviewed studies

Among the 87 reviewed articles, 23 were from 2003 to 2007 and 64 were from 2008 to 2013. The number for the latter period is more than 2.5 times that of the former period. This indicates an increasing interest on this topic during the last six years. All of these articles were published in 43 peer-reviewed journals. More than half of the articles were published in nine journals: Computers & Education (14 articles), the Journal of Computer Assisted Learning (7 articles), the Journal of Educational Computing Research (4 articles), Reading and Writing (4 articles), the Australasian Journal of Early Childhood (3 articles), the British Journal of Educational Technology (3 articles), the Journal of Research in Reading (3 articles), Reading Research Quarterly (3 articles), and Research in Developmental Disabilities (3 articles).

Table 1 provides an overview of the 94 studies in this review. The majority of the studies aimed to evaluate the influence of technologies on children’s learning, compare the effectiveness of interventions with and without technology-assisted learning, and investigate the roles of adults in children’s use of technology. Moreover, “technology evaluation,” “adults’ roles,” and “teaching approaches” are three emerging research topics during 2008-2013. In addition, more than half of the studies employed either experimental or quasi-experimental design, while more than a quarter of them were case studies. Lastly, about one-third of the studies concerned the learning of children from diverse backgrounds, including those from immigrant or low socioeconomic status families (13 studies) or had special needs (19 studies). This finding implies that researchers have noticed the affordances of technology for educators to scaffold the learning of these children.

| Table 1. Research purposes, research design, and research participants by publication year |
|-------------------------------------------------|-----------------|-----------------|-----------------|
| Publication Year                                | 2003-2007 n (%)| 2008-2013 n (%)| Total n (%)   |
| Research Purpose                                |                 |                 |                |
| Technology Evaluation                           | 9 (9.6)         | 26 (27.7)       | 35 (37.2)      |
| Technology vs. Tradition                        | 13 (13.8)       | 12 (12.8)       | 25 (26.6)      |
| Teaching Approaches                             | 0 (0.0)         | 7 (7.4)         | 7 (7.4)        |
| Mechanism Design                                | 2 (2.1)         | 4 (4.3)         | 6 (6.4)        |
| Adults’ roles                                   | 1 (1.1)         | 15 (16.0)       | 16 (17.0)      |
| Consumers vs. Creators                          | 1 (1.1)         | 4 (4.3)         | 5 (5.3)        |
| Research Design                                 |                 |                 |                |
| Experiment                                      | 4 (4.3)         | 24 (25.5)       | 28 (29.8)      |
| Quasi-experiment                                | 12 (12.8)       | 10 (10.6)       | 22 (23.4)      |
| One group pre/post-test or post-test design     | 2 (2.1)         | 2 (2.1)         | 4 (4.3)        |
| Case study                                      | 7 (7.4)         | 20 (21.3)       | 27 (28.7)      |
| Survey                                          | 1 (1.1)         | 8 (8.5)         | 9 (9.6)        |
| Others                                          | 0 (0.0)         | 4 (4.3)         | 4 (4.3)        |
| Research Participants                           |                 |                 |                |
| General                                         | 15 (16.0)       | 47 (50.0)       | 62 (66.0)      |
| Diverse                                         | 11 (11.7)       | 21 (22.3)       | 32 (34.0)      |

Cross-analysis results

Research purpose vs. research design

Table 2 shows the cross-analysis results of the research purposes and research designs. When the studies focused on describing how young children used technology in situ, they mostly adopted a case study design, followed by the survey method used to provide generalizable patterns of children’s technology use. It should be noted that these survey studies either collected data reported by adults (i.e., parents or teachers) or utilized data from nationwide databases. The experimental and quasi-experimental designs were employed by almost all studies aiming to compare
the effects of technology-supported learning with conventional approaches and those aiming to compare different technology designs (i.e., embedding teaching approaches and mechanism designs). Although experimental and quasi-experimental designs were also adopted by the majority of the studies that investigated the influence of adults’ roles in learning, studies that applied the case study method could provide in-depth observation of adults’ roles in children’s use of technology. With the least studied topic “consumers vs. creators,” all studies were conducted using the case study method.

### Table 2. Frequencies and percentages of research purposes and research designs

<table>
<thead>
<tr>
<th>Research Purpose</th>
<th>Experiment n (%)</th>
<th>Quasi-Experiment n (%)</th>
<th>One group pre/post-test or post-test design n (%)</th>
<th>Case study n (%)</th>
<th>Survey n (%)</th>
<th>Others n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology Evaluation</td>
<td>2 (2.1)</td>
<td>2 (2.1)</td>
<td>4 (4.3)</td>
<td>15 (16.0)</td>
<td>9 (9.6)</td>
<td>3 (3.2)</td>
</tr>
<tr>
<td>Technology vs. Tradition</td>
<td>8 (8.5)</td>
<td>15 (16.0)</td>
<td>0 (0.0)</td>
<td>1 (1.1)</td>
<td>0 (0.0)</td>
<td>1 (1.1)</td>
</tr>
<tr>
<td>Teaching Approaches</td>
<td>7 (7.4)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Mechanism Design</td>
<td>3 (3.2)</td>
<td>2 (2.1)</td>
<td>0 (0.0)</td>
<td>1 (1.1)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Adults’ roles</td>
<td>8 (8.5)</td>
<td>3 (3.2)</td>
<td>0 (0.0)</td>
<td>5 (5.3)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Consumers vs. Creators</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>5 (5.3)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

n: Total number of studies

We further applied a chi-square test and found the associations between research purpose and research design significant ($\chi^2 = 43.39, p < .001$). Researchers used the design of experiments (i.e., experimental, quasi-experimental, and one group pre/post-test or post-test design) more often than non-experiments (i.e., case study, survey, and others) when they compared technology-assisted and traditional ways of learning ($AR = 4.1$) or the effectiveness of technologies designed on the basis of different teaching approaches or mechanism ($AR = 2.7$). In contrast, the design of non-experiments outnumbered experiments significantly when the research purposes were “technology evaluation” ($AR = 5.2$) or “consumers vs. creators” ($AR = 2.7$).

### Developmental domains vs. learning effectiveness

Among 94 studies, 83 investigated the influence of technology on cognitive learning. Social learning is the second most frequent developmental domain emphasized by the studies ($n = 19$), followed by the emotional domain ($n = 12$) and the physical domain ($n = 2$). Most of the studies targeted one domain ($n = 76$), several studies considered the influences of technologies on two domains ($n = 15$) (e.g., cognitive and social domains or cognitive and emotional domains), and only three discussed results across more than three domains. Within the 83 studies that emphasized cognitive learning, different areas of learning were supported including language and literacy ($n = 47$), digital literacy ($n = 19$), math ($n = 12$), science ($n = 6$), and general cognitive abilities (e.g., problem solving, working memory, self-regulation, and creativity) ($n = 12$). In other words, cognitive learning is the most investigated domain, and under this domain, the field of language and literacy learning has attracted the most attention. The potentials of using technology to support learning in other domains and other cognitive fields are needed. As for learning results, 61 studies reported positive findings, 24 reported no differences, and only 2 reported negative outcomes. It seems that the majority of the studies indicated an overall positive effect of technologies on learning. However, it should be noted that two biases might contribute to the overwhelmingly positive results. One is that authors are more likely to submit papers with positive learning outcomes. The other is that the reviewed papers were limited to educational research. Papers in other fields (e.g., pediatric and psychology) may concern more about negative effects and are suggested to be included in future studies. Moreover, another important finding is that more than half of the studies ($n = 51$) found that the effect of technology-supported learning is conditional. An in-depth exploration of the conditional factors is elaborated in the following subection. Table 3 and Table 4 show the results of cross-analysis between developmental domains and learning effectiveness. Because a study with more than one result would be coded into different categories, only frequencies are provided to illustrate an overall distribution of developmental domains by learning effectiveness.
As mentioned earlier, we found that the influence of technology on children’s learning is conditional. We then further identified several conditional factors and categorized them into three key components: adults, children, and technology. Moreover, we proposed a typology that displays the complex relationships among the three components (see Figure 1). This typology shows the reciprocal interplay between children and technology, and adults play as mediators between children and technology. It should be noted that the effects of these factors are not conclusive. We did not attempt to synthesize these effects but highlighted potential avenues for future research.

A typology for conceptualizing young children’s learning with technology

As mentioned earlier, we found that the influence of technology on children’s learning is conditional. We then further identified several conditional factors and categorized them into three key components: adults, children, and technology. Moreover, we proposed a typology that displays the complex relationships among the three components (see Figure 1). This typology shows the reciprocal interplay between children and technology, and adults play as mediators between children and technology. It should be noted that the effects of these factors are not conclusive. We did not attempt to synthesize these effects but highlighted potential avenues for future research.

Figure 1. A typology of factors that influence children’s learning through technology.

The children aspect

The impact of technology on children’s learning is conditional by children’s age, experience, time spent using the technologies, and gender. In terms of children’s age,
intervention on children of different age groups, older children tended to outperform younger children (e.g., Andrews, Woodruff, MacKinnon, & Yoon, 2003; Cviko, McKenney, & Voogt, 2012; Korat & Shamir, 2012), or the intervention had a statistically significant effect on older children but not on younger children (e.g., Manches et al., 2010; Watson & Hempenstall, 2008)

Also, children’s experiences, including (a) prior knowledge and (b) computer access at home, were related to their technology-related learning. Children with greater prior knowledge made more progress in their performance after using technology. For example, Zacharia, Loizou, and Papaevripidou (2012) found that kindergarteners who had correct prior knowledge of a balance beam learned more from the experimentations regardless of the different ways in which they conducted the experiments. In contrast, a study suggested that having more prior knowledge may be a disadvantage for children’s technology-related learning. Levy (2009) revealed that young children’s accumulated knowledge in print reading in school may impede their development of reading multimodal computer texts. She found that children were confident using strategies to read computer texts before entering school. However, after learning in school that reading was “the decoding of print in school-based books” (p. 88), their confidence in reading computer texts decreased over time.

Another factor is children’s computer access at home. Studies have shown that children’s computer access at home is positively correlated with their performance in reading (e.g., Bittman, Rutherford, Brown, & Unsworth, 2011; Magnan & Ecalle, 2006), math (e.g., Judge, Puckett, & Bell, 2006), cognitive development (Fish et al., 2008), and computer-esteem and skills (Hatzigianni & Margetts, 2012; Sackes, Trundle, & Bell, 2011). However, Puckett and Bell (2006) found that children’s frequent use of computer programs for reading was negatively correlated with their reading achievement. They suggested that young children with low reading achievement would benefit more from teachers’ reading instruction than from computer programs for reading.

Time is another factor that influences children’s learning with technology. Time factors include (a) the treatment retention effects and (b) the time that children spent using technologies. For example, Kyle et al. (2013) investigated the retention effect of a computer-assisted reading intervention and found children maintained their gains on reading skills four months after the intervention. Also, studies have revealed that the more time that children spent on the technology-supported intervention, the better performance they had (Lonigan et al., 2003; Segers & Verhoeven, 2005). However, Ryokai, Farzin, Kaltman, and Niemeyer (2013) found that some children in their study showed negative performance slopes regarding the ability to track multiple objects over computer-assisted intervention sessions.

Gender difference in computer-related performance is another topic that researchers emphasized. The effect of technology use for boys and girls varied. A study indicated that girls had a higher developmental rate of computer skills than boys (Sackes et al., 2011), while another study found that boys had better learning outcomes than girls after playing a noncompetitive mathematics game (Wei & Hendrix, 2009).

Other children’s backgrounds, such as second language learners and children’s social economic status, are factors that researchers examined when investigating topics regarding how technology influences children’s learning (e.g., Kim & Chang, 2010; Sackes et al., 2011).

**The adults aspect**

The role of adults is another critical component in the typology for conceptualizing the interplay between children’s learning and their technology use. We discuss the four aspects of adults’ roles in children’s technology use as follows. First, adults facilitate children’s engagement in using technology to learn (e.g., Eagle, 2012; McKenney & Voogt, 2009). Children learned more from using technology when adults provided them with a safe climate, encouraged them to participate in conversation, involved them in establishing the goals of the activity, and maintained their interaction with the adults and with the technology.

Second, adults adapt their teaching to respond to children’s learning with technology. For example, Clements and Sarama (2008) stated that children gained most from a technology-integrated, mathematics curriculum in which the teachers adapted the learning activities based on their understanding of the learning trajectories and the children’s prior knowledge. Similarly, Shamir, Korat and Barbi (2008) found that teachers’ adaptive instruction for technology-
assisted learning would lead students to better learning outcomes. In other words, when teachers instructed children to read e-books with their peers, the children showed more improvement in their reading skills than those who read e-books alone.

Third, adults’ perceptions of children’s technology use influence how they support or do not support their children’s learning through technology. Some adults had a positive attitude toward technology and made efforts to integrate technologies into their curriculum or to engage children in technology-related activities (e.g., Cviko et al., 2012; Fessakis, Gouli, & Mavroudi, 2013). In contrast, in Wolfe and Flewitt’s (2010) study, most of the participating parents and teachers were concerned that children’s frequent use of technology may impede their development, and these adults either restricted the amount of time their children could spend using computers or did not encourage or facilitate children’s technology use.

Fourth, adults’ teaching in conjunction with technology-assisted learning maximizes the effect of technology on children’s learning, whereas adults’ teaching and technology-assisted learning alone had less effect on children’s learning gains (Eagle, 2012; Segal-Drori, Korat, Shamir, & Klein, 2010). Segal-Drori et al. (2010) found that children who read electronic books and received teachers’ instruction that promoted emergent reading outperformed those children who read electronic books without adult instruction. They also outperformed those children who read printed books with adult instruction.

The technology aspect

Three technology-related factors are discussed: the mechanism design of the technologies, the teaching and learning approaches applied to the design of the technologies, and the content of the technologies.

Studies have revealed the impact of different mechanism designs of a technology on children’s behavior (e.g., Druin et al., 2003; Korat & Shamir, 2012). For instance, Druin et al. (2003) found that different designs of interfaces supported different aspects of children’s collaborative experiences. In their study, they developed two collaborative conditions for paired children to work on computer tasks. In one condition, each of the paired children used an individual mouse and controlled the interface independently; in the other condition, each action needed to be confirmed by the other paired child. The authors found that the former, flexible condition led children to talk more about the content of the tasks. The latter, structured condition led the children to talk more about the shared goals and function of the mouse.

In terms of applying different teaching and learning approaches in the design of the technologies, researchers have examined the effects of computer programs or games developed on the basis of different teaching and learning approaches. Examples of these approaches are: systematic phonics training versus unsystematic phonics training (De Graaff et al., 2009), synthetic phonics intervention versus analytic phonics intervention (Comaskey, Savage, & Abrami, 2009), bottom up literacy training versus top down literacy training (Helland, Tjus, Hovden, Ofte, & Heimann, 2011), whole language games versus phonics skills games (Segers & Verhoeven, 2005), a prediction-observation-explanation (POE) model science game versus a non-POE model science game (Hsu, Tsai, & Liang, 2011), and approximate number training versus exact number training (Obersteiner, Reiss, & Ufer, 2013).

Lastly, the content of the technologies also plays a role in children’s development. Conners-Burrow, Mckelvey, and Fussell (2011) indicated that children’s viewing of age-inappropriate content of television programs (i.e., PG-13 or R-rated videos and movies) was related to their aggressive and hyperactive behavior problems in the classroom. They stressed that it was the content of the television programs rather than the amount of time children spent watching television that had an impact on the development of their social skills.

Emergent discussion on digital literacy

Among the research topics devoted to the understanding of children’s cognitive development and technology use, children’s literacy learning was the most studied. While most researchers emphasized such traditional ideas of literacy, some have explored new forms of literacy made possible by the development of digital technologies and the
internet: digital literacy. These researchers examined young children’s digital literacy development from two aspects: (a) skills and competences needed to use technology, and (b) perceptions of technology use.

**Skills and competences needed to use technology**

The majority of the studies related to young children’s digital literacy development involved the skills and competences that children had or needed to use technology including:

- Children’s ability to use a mouse and touchscreen: Studies have been conducted to examine the ways to improve young children’s computer skills with a mouse or touchscreen. For example, Shimizu, Yoon and McDonough (2010) developed a training program that was found to be effective in reinforcing developmentally-delayed preschoolers’ point-and-click skills with a mouse and their eye-hand coordination. Donker and Reitsma (2007) compared two ways of using a mouse, drag-and-drop versus click-move-click, and found that the former resulted in fewer interaction errors and was more appropriate for kindergarteners and first graders.

- Children’s learning of computer programming (Fessakis et al., 2013; Matthews & Seow, 2007): Studies have revealed that young children can learn the basic concepts of programming. In Fessakis, Gouli and Mavroudi’s (2013) study, the participating kindergarteners were able to use commands to solve programming problems. Through the intervention program, kindergarteners learned basic programming concepts, including “command, sequential execution of commands, program, logical errors, testing and debugging of programs” (p. 94).

- The digital tools and technology-related literacy practices that children had at home (Plowman, Stephen, & McPake, 2010; Plowman et al., 2011): Researchers adopted a sociocultural approach to literacy learning and found that preschoolers encountered technology and read digital texts in their daily familial life. These children knew how to play DVDs, use remote controls to locate and change TV channels, use digital cameras, mobile phones or tablets to take pictures, resize windows on a computer, type letters and use drawing tools on computers, play games on computers, mobile phones or tablets, and so on. Children had accumulated the knowledge of digital literacy from daily technology-related practices at home. Researchers are concerned with the discrepancy between home and school regarding children’s learning of digital literacy. This is because teachers in preschools often do not draw upon children’s knowledge of digital literacy in order to advance their learning with technology (Plowman et al., 2010).

- Children as creators of multimodal, digital texts: McPake, Plowman, and Stephen (2013) described that a three-year-old boy in the U.S. was proficient in taking photos. He and his five-year-old sibling sent photos and emoticons when making video calls to their relatives in Australia. The researchers stated that the boy learned to develop a storyline “in visually meaningful and engaging ways, beginning to develop the skills to create new and (socially) valuable narratives” (p. 427). In O’Mara and Laidlaw’s (2011) study, young children were able to create multimodal stories consisting of audio narratives, pictures, video clips, and digital drawings and texts. Furthermore, they had the ability to transform content from one mode (sign system) to another to make sense of and communicate with multimodal texts. Mills (2011) indicated that this transformation process, called transmediation, was demonstrated by eight-year-olds in her study. The children translated written texts into images and their drawings into a film. They also transformed their handwritten comics into digital online comics. The role of children as creators of technology will be discussed in a later subsection.

- Children’s ability to read and comprehend multimodal, digital texts: Studies have revealed that young children are able to use multimodal cues to comprehend the meaning within the contexts of digital texts. Such multimodal cues include pictures, symbols, sound, images, and gestures, which were used across a variety of technologies, such as TV, computers, mobile tablets, mobile phones, game consoles, and touch screens (e.g., Levy, 2009; McPake et al., 2013).

- Children’s ability to search for information online: Young children are capable of using search engines, such as Google and YouTube to locate information they need. For example, Davison (2009) described that a six-and-half-year-old boy used Google to search for more information about the lizard he learned from a book. O’Mara and Laidlaw (2011) found that 3- and 5-year-olds were interested in selecting and viewing online programs and video clips on YouTube or other internet resources.
Perceptions of technology use

The other aspect of children’s digital literacy development is their perceptions of technology use, including (a) perceptions of the social and cultural roles of technologies, and (b) perceptions of their capability to use computers, that is, their computer self-esteem. Studies have shown that preschoolers had learned the social purposes of technology (McPake et al., 2013; Plowman et al., 2012). These social purposes included communication, maintaining social ties, entertainment, study, and adults’ employment. With regard to children’s learning of the cultural conventions of using technology, McPake et al. (2013) described that when making a video call, a three-year-old boy considered his communication partners’ viewpoints and emotions, and then he chose the most appropriate photos to send to them. His behavior demonstrated that he was aware of the culturally appropriate ways of making a video call. In terms of children’s development of their computer self-esteem, it was positively correlated with their computer access at home and in school (Hatzigianni & Margetts, 2012).

Technology supports children’s social development

Some researchers believe that children at a young age should avoid using technology because it keeps them from interacting with others and therefore impedes the development of their social skills (e.g., Armstrong & Casement, 2000; Cordes & Miller, 2000). In contrast to this viewpoint, we found that only one reviewed study revealed that the use of a touchscreen increased children’s behavior of pursuing individual goals instead of collaboratively achieving the same goal, implying that the use of technology may hinder children’s social development (Romeo, Edwards, McNamara, Walker, & Ziguras, 2003). Most of the studies, however, showed that various technologies support children’s social development (see Table 3).

Young children’s social development was supported by technologies in three aspects. First, a variety of technologies enhanced children’s collaboration and interaction with peers. For example, Infante et al. (2010) found that a video game, which was designed for multiple players to use one computer screen and several input devices, encouraged kindergarteners to collaborate and communicate in order to complete the game tasks. Lim (2012) examined kindergarteners’ social behavior in the computer area in a classroom. The author argued that in that area, the kindergarteners learned useful information and engaged in learning through active interaction with their peers.

Second, the technologies used at home facilitated adult-child interaction and maintained family relationships. Researchers described how young children worked with adults (e.g., parents, grandparents, relatives) to achieve the shared goal of the technology-related activity and reinforce their ties with family members. For instance, the 3-6-year-old grandchildren and the grandparents helped each other in computer activities. While the children taught their grandparents how to play a computer game, the grandparents helped the children with the linguistic and cultural knowledge needed to play the game (Kenner, Ruby, Jessel, Gregory, & Arju, 2008).

Third, technology was related to young children’s development of multiculturalism. As Perry and Moses (2011) indicated, television programs in the U.S. reinforced immigrant children’s development of their cultural identification with their country of origin, Sudan. However, in terms of the effect of the television programs in changing young children’s attitudes toward the non-majority, Persson and Musher-Eizenman (2003) indicated that regardless of how the show was presented (i.e., real people, animation, puppets), after watching the show, the 3- to 6-year-olds in the U.S. maintained their preference for White over people of other ethnic groups.

Children’s roles as consumers and creators of technologies

As shown in Table 1, unlike most of the studies that viewed children as consumers of pre-designed learning content delivered by technologies, only five studies explored and observed children’s use of technology as creators, either by creating digital artifacts such as multimodal texts, paints, photos, and videos (e.g., McPake et al., 2013; Mills, 2011) or by writing or modifying programs to solve problems (Fessakis et al., 2013). The roles of children in using technologies can correspond to the tutor/tool/tutee framework proposed by Taylor (1980). This framework provides an important foundation to classify all educational computing when researchers consider the roles of technologies in education. When children receive instruction like consumers, technologies are employed as tutors for their learning. Most of the technologies adopted in the reviewed studies played this role. For example, a computer game was
designed to teach children about light and shadow (Hsu et al., 2011) or e-books were implemented to promote children’s literacy (e.g., Korat, 2010). When children become creators of learning, the technologies can be either used as tools or tutees. For example, tablet PCs, digital cameras or recorders, and computer programs are all tools for children to express and communicate their ideas or to engage in social interactions with family members. With the increasing availability of various technologies at home and in schools, there would be a great opportunity for promoting children’s learning by using these technology tools. On the other hand, the Logo-based programing environments, the Ladybug leaf and the Ladybug maze, adopted by Fessakis and his colleagues (2013) demonstrated an example of using technology as tutees. Children can modify the program or create new programs to lead the ladybug (the character in the software) through a leaf. Scholars have suggested that by using technology as tutees, learners could learn in more depth, learn more about the learning process, and link their experience to the fundamental concepts of learning subjects (c.f., Taylor, 1980).

Implications

Several implications for future research are proposed. The typology proposed in this study is expected to contribute to both research and practice. Following this typology, different factors influencing the results of technology-supported learning in young children can be systematically investigated and discussed. On the one hand, with more studies examining the children, adults, and technology factors, this typology can be validated further. The typology can also help educators and technology designers to consider the factors related to the learning outcomes. How to apply this typology to design or evaluate a technology-supported instruction or system will need to be discussed in the future. Another emerging concern of this reviewed topic is the roles of children as consumers or creators of technologies. How technologies can support young children to become creators will rely on future design of technologies and evaluation of their implementation. Moreover, how different types of technology can facilitate various developmental domains of children will need future exploration. It should be noted that although 13 studies provided the learning theories or teaching strategies they adopted to design the technologies, most of the reviewed articles did not explicitly state the learning theories underlying their studies. More studies framed on the basis of established learning theories are needed. Additionally, the learning results in this study were roughly classified to positive, negative, no difference, and “depends.” To examine the learning effectiveness more precisely, a statistical meta-analysis is suggested to examine the degree of effectiveness of the learning results. Lastly, how different types of measures may influence the results of learning outcomes will need further investigation.

References


Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. Learning and Instruction, 23, 125-135. doi: 10.1016/j.learninstruc.2012.08.004


Learning of Abstract Concepts through Full-Body Interaction: A Systematic Review

Laura Malinverni and Narcis Pares
University Pompeu Fabra, c. Roc Boronat 138, 08018 Barcelona // University Pompeu Fabra, c. Roc Boronat 138, 08018 Barcelona // laura.malinverni@upf.edu // narcis.pares@upf.edu

*Corresponding author

ABSTRACT

Over the past ten years several learning environments based on novel interaction modalities have been developed. Within this field, Full-body Interaction Learning Environments open promising possibilities given their capacity to involve the users at different levels, such as sensorimotor experience, cognitive aspects and affective factors. However, Full-body Interaction is still a young field and research on design and assessment methods offers a fragmented panorama from which it is not possible to derive clear research solutions. Starting from this necessity, we present a systematic review of educational applications developed as Full-body Interaction and the results of their empirical evaluation. Our analysis offers instruments to systematize the multiple aspects involved in the design and assessment of Full-body Interaction Learning Environments and provides guidelines for novel research paths.

Keywords

Full-body interaction, Whole-body interaction, Embodied interaction, Technology-enhanced learning

Introduction

For several years technology-enhanced learning has mainly focused on the content of the interaction, relegating the physical interface and the interaction modality to a pure instrumental role. However, over the past years, under the term of “Embodied Interaction” - defined as “the use of the physical world to interact with digital technologies” (Dourish, 2004) - several Learning Environments based on novel interaction modalities have been developed. Promising possibilities can be found in Full-body Interaction, understood as using the movements and the actions performed in the physical space by the body of the user, as mediators of the interactive experience. From a theoretical perspective Full-body Interaction has the potential to support learning (Revelle, 2013) by involving users at different levels such as sensorimotor experience, cognitive aspects and affective factors.

In this context an increasing number of Full-body Interaction Learning Environments (FUBILEs) have been designed to support learning. However, due to its novelty, findings related to its evaluation do not allow identifying effective and generalizable design strategies. This situation offers a fragmented panorama from which meaningful conclusions about design choices may not be deduced.

Starting from this necessity, we present a systematic review of educational applications developed as Full-body Interaction Learning Environments and aimed at fostering the learning of abstract concepts. A total of thirty selected papers, published during the last ten years, have been reviewed. The purpose of the review is to provide a clear systematization of theoretical approaches, design strategies and evaluation methods for FUBILEs. The analysis shows the effectiveness of FUBILEs in engaging learners and in grounding abstract concepts. Furthermore, it allows us to formalize guidelines for designers and paths for future research especially in the fields of design, evaluation methods and comparative studies.

Method of analysis

This article reviews studies published in the last ten years (2003-2013) related to the use of Full-body Interaction for learning of abstract concepts. The literature reviewed proceeds from peer-reviewed studies published in English in scientific journals, proceedings of international conferences, symposia and book chapters. The search was carried out through the consultation of the following online databases of academic resources: ERIC, JSTOR, PapersFirst, ACM, IEEE, WilsonWeb, Elsevier, InformaWorld, Mary Ann Liebert, SpringerLink, Wiley Interscience, MIT Press and SAGE.
The initial research was based on searching the terms “learning” and “education” combined with the keywords: embodied interaction, full-body interaction, whole-body interaction, motion-based interaction, gesture-based interaction, bodily interaction, and kinesthetic interaction. We then searched for articles cited in the initially found papers.

The selection was based on the inclusion criteria of papers that describe learning environments designed for neurotypical populations, which address knowledge acquisition as the primary goal and that are based on Full-body Interaction. Theoretical papers and studies on other forms of embodied interaction, robotics and construction kits were excluded.

Finally, we have included a total of thirty papers published between 2003 and 2013 (see Table 1). Since research in FUBILE is an emerging field, the chronological distribution of publications is strongly skewed toward recent years, with twenty-one papers being published between 2011 and 2013 and only ten published between 2003 and 2011. Also, as a consequence of the novelty of the field, most papers proceed from international conferences mainly dedicated to HCI or Learning Science. Only five of the reviewed papers proceed from journals and three are book chapters.

We oriented the analysis toward providing an exhaustive panorama of FUBILEs to offer a clear systematization of theoretical frameworks, design strategies and evaluation methods. For this purpose, the review focuses on the following aspects: the underlying theoretical frameworks; the design methodologies used; the interaction design choices; the evaluation methods applied; and the outcomes of the empirical analysis.

We used content analysis to classify the defined research topics into categories. Our coding procedure was realized through Nvivo software and comprised a blend of a priori categories with new categories emerging from a grounded analysis.

<table>
<thead>
<tr>
<th>Project</th>
<th>Theoretical Framework</th>
<th>Design Strategy</th>
<th>Educational Context</th>
<th>Interaction Design</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Adachi et al., 2013)</td>
<td>Human Sugoroku</td>
<td>X</td>
<td>X</td>
<td>Science</td>
<td>Children</td>
</tr>
<tr>
<td>(Anastopoulou et al., 2011)</td>
<td>Kinematics graphic</td>
<td>X</td>
<td>X</td>
<td>Physics</td>
<td>Adults</td>
</tr>
<tr>
<td>(Antle et al., 2008)</td>
<td>SoundMaker</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Music</td>
</tr>
<tr>
<td>(Antle et al., 2013)</td>
<td>SpringBoard</td>
<td>X</td>
<td>X</td>
<td>Social</td>
<td>Adults</td>
</tr>
<tr>
<td>(Carreras &amp; Parés, 2009)</td>
<td>(Carreras &amp; Parés, 2009)</td>
<td>X</td>
<td>X</td>
<td>Science</td>
<td>Children</td>
</tr>
<tr>
<td>(Charoenying et al., 2012)</td>
<td>(Charoenying et al., 2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(Cress et al., 2010)</td>
<td>(Cress et al., 2010)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Math</td>
</tr>
<tr>
<td>(Edge et al., 2013)</td>
<td>(Edge et al., 2013)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Language</td>
</tr>
<tr>
<td>SpatialEase (Gronbæk et al., 2007)</td>
<td>SpatialEase (Gronbæk et al., 2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Physics</td>
</tr>
<tr>
<td>IGameFloor (StepStones)</td>
<td>IGameFloor (StepStones)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Various</td>
</tr>
<tr>
<td>(Hashagen et</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Music</td>
<td>Adults</td>
</tr>
</tbody>
</table>
The theoretical framework

Research in FUBILEs is mainly grounded on the benefits that physicality may provide to facilitate learning and enhance user experience. This idea is supported by different theoretical frameworks, which range from pedagogy to cognitive science and physiology. From our analysis, we identified three main approaches: (1) based on developmental psychology and pedagogical theories, (2) relying on Embodied Cognition and (3) based on the physiological benefits that exertion can bring to cognition. To situate the reviewed projects, we provide an overview of the main theoretical frameworks. Despite these approaches are interrelated, we describe them as separate sections for the sake of clarity.

The developmental psychology framework

From a pedagogical perspective FUBILEs are mainly based on constructivist and constructionist frameworks. The central pedagogical value is posed on the idea of learning-by-doing, understood as the fundamental role of hands-on activities and active experiences in the learning process.
The works of Jean Piaget, Seymour Papert and Jerome Bruner support such theoretical approach. Even when these authors have different perspectives on child development, they all agree that knowledge emerges as a result of people’s action-in-the-world (Ackermann, 2001). Such perspective supports the action-oriented approach of Full-body Interaction and suggests its potential for facilitating the construction of knowledge through the internalization of actions.

A particular relevance is given to Piaget’s notion of schemata, Papert’s notion of body-syntonicity and Bruner’s concept of enactive representation. Piaget suggests that the cognitive structuring of children is partially based on the extension of physical schemata, which represent internalized patterns of activity that are then used for thinking (Antle et al., 2008). Papert (1980), through the notion of body-syntonicity, describes instructional designs based on using children’s knowledge about their own bodies to stress the ‘resonance’ between abstract concepts and what people know about themselves (Watt, 1998). Finally Bruner’s theory about modes of representation of knowledge suggests that providing learners with different ways of thinking (enactive, iconic, symbolic) can facilitate the learning process (Di Paolo et al., 1991).

The embodied cognition framework

Twenty-one papers base their designs on the Embodied Cognition framework. Embodied Cognition has its philosophical roots in Merleau-Ponty who, by defeating Cartesian dualism, re-introduces the body as a skilled subjectivity that helps knowledge construction (Shusterman, 2013). Such framework emerged in cognitive science around fifty years ago and has been incorporated in HCI during the last two decades (Antle, 2013). According to this framework, almost all cognitive processes are influenced by physical states, bodily structures (Wilson, 2002) and experiential opportunities.

Embodied Cognition, by focusing on the fundamental role of action and perception in shaping cognitive processes, coherently relates with the pedagogical approach of learning-by-doing and provides a scientific ground for defining design strategies for FUBILEs. With this context, particular relevance is given to Barsalou’s Grounded Cognition approach, Johnson and Lakoff’s theory on the embodied nature of linguistic concepts, and to the studies on the relation between gesture and thought.

Barsalou’s Grounded Cognition suggests that mental representations are grounded in motor areas of the cortex: when knowledge is needed, the perceptual and motor states acquired during experience are reactivated through simulation (Barsalou, 2008). The findings of Lakoff and Gallese support this hypothesis suggesting that the formation of linguistic concepts meet its bases in the sensorimotor system (Gallese & Lakoff, 2005). Similarly, Johnson proposes the Embodied Metaphor Theory, which suggests that abstract concepts and conceptual metaphors are based on image schemas that derive from physical actions (Johnson, 1987). Finally, the studies carried out by Goldin-Meadow (2011) show the intertwined relation existing between gestures, language and learning outcomes and propose instructional models capable of taking into account the role of body in thinking processes. According to her research, gestures can predict and provoke learning by displaying knowledge that cannot be expressed verbally yet and by facilitating knowledge construction.

The physiological framework

The third theoretical approach is related to the impact of physical activity on cognitive functioning, memory, attention allocation and academic performance (Castelli et al., 2007; Hillman et al., 2008). Recent researches show that aerobic exercise can improve several aspects of cognition, suggesting a physiological relation between physical activity and academic success. However, despite the raising importance of studies relating physical activity with academic success (Raine et al., 2013) it is interesting to notice that only three of the reviewed papers encompass such aspect in their design framework.
Design strategies

The overview of the theoretical frameworks shows that research in FUBILEs is based on findings that are consistent with research in developmental psychology, cognitive science and physiology. Developmental psychology and Embodied Cognition, by stressing the tight relation between body and knowledge construction, suggest the possibility of using relevant bodily actions to facilitate the grounding of abstract concepts. The studies on the physiological effects of exertion suggest that FUBILEs can be beneficial from a perspective that encompasses also the cognitive process at the very ground of learning such as memory and attention.

However, since HCI is an applied science it is necessary to analyze how these findings can be used to inform design. Such task represents a challenging requirement that implies concretizing complex theoretical frameworks into specific instances. To observe of how this challenge has been addressed, the following sections describe the most relevant design strategies and their relation with the underlying theoretical frameworks.

Only three papers use a design method that involves stakeholders or end-users in the design process (Enyedy et al., 2012; Grønbæk et al., 2007; Johnson-glenberg et al., 2010). Namely two of them are based on the collaboration with teachers (Enyedy et al., 2012; Johnson-glenberg et al., 2010) and one involves both teachers and children through the use of participatory design methods (Grønbæk et al., 2007). The rest of projects mainly arise from a perspective grounded on exploring theoretical constructs that, in some cases, are combined with empirical validation through user testing.

In this context three main design strategies were identified: a semantic approach, an approach based on the adaptation of existing materials and a physiology-oriented approach.

The semantic approach

Most analyzed projects rely on a semantic approach. This implies a design strategy oriented toward using the actions, events and activities of FUBILE as a reference for constructing meaning. This approach is generally aimed toward facilitating an embodied experience of a certain concept or toward representing an abstract concept as a concrete instance. From a theoretical perspective it is related to projects based on the Embodied Cognition framework.

Within this context, different semiotic resources, understood as “resources for making meaning” (Van Leeuwen, 2004) have been utilized to communicate the target contents. The greatest relevance is given to the role of actions as a vehicle to transmit meaning. Examples can be found in the Method for Meaning Generation proposed by Carreras & Pares (2009) and in the Embodied Metaphor Approach used by Antle (2013) and Holland (2011).

The Method for Meaning Generation focuses on operationalizing actions in terms of attitudes that we want the users to adopt and concentrates the design strategy on defining how the user will physically interact with the system. Examples can be found in “Connections” (Carreras & Pares, 2009) and “WaterGames” (Pares et al., 2005) where the values related to certain behavior (e.g., users hold hands with each other) serves as a link to express a specific concept (e.g., collaboration between scientists).

The Embodied Metaphor Approach proposes the use of Johnson’s theory of conceptual metaphor (1987) to inform design. An example can be found in “Springboard”, where the concept of “the balance of social justice” is derived from the physical experience of balancing our own bodies (Antle et al., 2013). Other examples are “Song Walker Harmony Space” and “Sound Maker” where embodied metaphors are applied to musical concepts (Holland et al., 2011).

Similarly to actions, another semiotic resource is space. An example can be found in Cress et al. (2010) where the spatial format supports children’s understanding of numerical comparison through the use of spatial cues (e.g., left = smaller, right = bigger). At the same time, the three projects based on Johnson’s Embodied Metaphors, use both physical and spatial metaphors as resources in their design. Interestingly, Antle (2008) points out that in “Sound Maker” children tended to use more spatial elements than bodily-based elements as sources for meaning-making. Another project based on the features of space is “Arquimedes” (Malinverni et al., 2012), where the physical
affordances of the interface are used to communicate meaning. The project, based on the use of a large inflatable slide, employs the gravity experienced on the sliding surface to communicate concepts related to gravity itself.

Other projects propose the coupling between physical movement and computational feedback as the main semiotic resource for meaning-making (Anastopoulou et al., 2011; Charoenying et al., 2012). Both projects focus on the learning of graphical representations of mathematical and physical concepts and stress the role of mapping between changes in the user’s movements and changes in graphs as the mediator of learning. For example “Bar Graph Bouncer” (Charoenying et al., 2012) couples the number of jumps performed by children with bar graph representations of quantities.

Finally, the project “PUSH” (Johnson-glenberg et al., 2011) and “Climate change” (Lyons et al., 2012) use the notion of effort as a semiotic resource. The former project uses the effort experimented by children while pushing virtual objects to address concepts of Newtonian forces. The latter uses the notion of effort to allow children understand the severity of climate change.

Adapting existing materials

Eight projects have a design strategy based on adapting existing materials to Full-body Interaction. Different strategies have been identified: (1) the adaptation of already existing applications, (2) the inspiration from traditional physical games and (3) the adaptation of educational material.

The project “Human Soguroku” adapts an already existing application designed for a desktop computer to Full-body Interaction for making the experience more immersive (Adachi et al., 2013). The project “SpatialEase,” instead draws inspiration from an existing application and from the traditional physical game of “Simon says” (Edge et al., 2013).

The use of traditional games for developing FUBILEs finds examples in the projects “IGameFloor” (Grønbæk et al., 2007), “HOPSCOTCH” (Lucht & Steffi, 2013) and “WaterGames” (Pares et al., 2005). In “IGameFloor” and “HOPSCOTCH” the mechanics of the traditional games “Twister” and “hopscotch” are used in didactic exercises. Quite differently “WaterGames”, instead of embedding novel content into already existing gameplay, looks for an existing game that shares a conceptual affinity with the addressed topic and therefore works as a semantic reference (e.g., “Ring-a-ring Roses” associated to “respect for diversity”) (Pares et al., 2005).

Finally, the project proposed by Lee (2012), adapts the educational material of “Live your life in English” to a FUBILE to instill the naturalistic approach of English conversations and provide a context for authentic learning.

No explicit relation has been identified between a specific theoretical framework and these design strategies. Adapting existing materials can have its pros and cons. On one hand, it facilitates understanding through the use of culturally established models. On the other, especially in those cases based on adapting existing desktop applications, it could run the risk of reducing the potential of Full-body Interaction to that of merely emulating mouse-based interaction paradigms with the body as a large-scale pointing device.

Physiology oriented approach

Two papers propose a design strategy based on a physiological approach and design interaction according to the relation between physical and cognitive workloads (Kiili et al., 2012; Lyons et al., 2012). This approach correlates with studies on the impact of physical activity on cognitive functioning, memory and academic performance. Such strategy focuses on analyzing the tight relation between physical arousal, attention and memory formation. It therefore suggests the necessity of balancing the physical workload with the cognitive load and proposes designs aimed toward offering an amount of physical activity that is beneficial for learning processes.
The educational context

Learning goals

Eighteen projects refer to STEM topics: seven refer to mathematics, seven to physics and four to science. This predominance can be associated with the governmental efforts to foster STEM and with the affordances provided by Full-body Interaction. Studies on mathematical cognition suggest the embodied and spatial nature of mathematical concepts (Lakoff & Núñez, 2000), and several studies suggest a correlation between spatial thinking skills and STEM academic performance (Newcombe, 2010).

Four projects focus on learning abstract musical concepts and three on learning a second language. Projects that address learning musical concepts (Antle et al., 2008; Holland et al., 2011; Volpe et al., 2012) arise from the idea of using physicality and spatiality to provide a concrete experience of abstract concepts such as tonal harmony. Projects about language learning arise either from the necessity of providing authentic learning experiences to the users (Lee et al., 2012) or from already existing instructional methods such as the Total Physical Response, which addresses second language learning through the co-production of spoken commands with bodily action (Edge et al., 2013).

Five of the analyzed projects focus on social sciences and particularly on topics that require a moral engagement such as social justice (Antle et al., 2013), environmental issues (Lyons et al., 2012), cultural diversity (Lim et al., 2011) or respect for diversity (Pares et al., 2005). The selection of these topics find their rationale in the capacity of direct experience to foster emotional arousal, suggesting the hypothesis that FUBILE could be capable of producing a higher impact, evoking empathy and producing behavioral changes.

Finally, three projects are not designed for a specific learning goal but work as frameworks where different contents can be implemented (Gronbæk et al., 2007; Kiili et al., 2012; Lucht & Steffi, 2013). Such a general approach correlates with the adoption of the theoretical framework based on the physiological benefits of exertion.

According to the revised version of Bloom’s taxonomy, learning goals can be structured into four classes of knowledge: factual knowledge (recall of elements), conceptual knowledge (knowledge of the core concepts), procedural knowledge (applied knowledge about “how to” do something) and metacognitive knowledge (knowledge about one's own cognition) (Krathwohl, 2002). Regarding these classes, projects based on frameworks where different contents can be implemented mainly address factual knowledge through didactic exercises (e.g., select the right answer). Instead, projects grounded on Embodied Cognition mainly address conceptual and procedural knowledge.

Target users

Out of the thirty projects, twenty are designed for children, four for teenagers and seven for adults. In projects designed for children the target age ranges from preschoolers (Cress et al., 2010) to primary school, with a high predominance of children between 10 and 12 years old. A possible explanation of this distribution can be found in children’s developmental trajectory; i.e., children around 11 years old are in a stage in which they start to think abstractly, reason about hypothetical problems (Perinat & Laluzea, 2007) and have well developed motor capabilities.

Context of use

Twelve projects arise from laboratory research and are tested in experimental settings. Other twelve are designed to be embedded in a school environment. Five are for museum spaces and one for a public space. The difference between settings has a fundamental role in defining design requirements, determining the educational experience and the empirical evaluation. This is clearly seen by comparing the educational experiences in an experimental setting or in a school environment. In most projects tested in the laboratory, participants use the system once for a limited amount of time and without any other educational support. Instead, in projects such as “Learning Physics” (Enyedy et al., 2012) the system is embedded in a complete educational program lasting fifteen weeks.
Interaction design

Physical configuration

Due to the physical and spatial nature of FUBILEs, a first aspect to consider is the configuration of the physical interface. Most projects are based on audiovisual outputs that combine visual representations and audio effects. Twenty-nine are mainly visual, while one uses only audio output (Antle et al., 2008).

In seventeen projects the physical interface is based on the use of vertical screens or wall projections, requiring users to interact in the space in front of the display. Within this configuration few differential elements are present, such as adding interactive sensors in the adjacent space (Holland et al., 2011; Lucht & Steffi, 2013; Lyons et al., 2012). On the other hand, twelve projects are based on floor projections, which allow the user to move around the periphery or directly on the visual output (Carreras & Pares, 2009; Cress et al., 2010; Grønbæk et al., 2007; Hashagen et al., 2009; Johnson-glenberg et al., 2010; Lindgren & Moshell, 2013).

Interestingly, only two projects present a substantial difference in the physical interface: “Arquimedes” (Malinverni et al., 2012) and “WaterGames” (Pares et al., 2005). In “Arquimedes”, the physical interface is a large inflatable slide, with a virtual environment projected on the sliding surface. In “WaterGames”, the physical interface is a group of water fountains, where the output is constituted by water springs.

The homogeneity in the physical interfaces should provoke a reflection on the affordances of displays such as vertical screens and floor projections and suggest possible paths for innovation.

Single user or multiple users

Eleven projects are designed for a single user, whereas nineteen are designed for multiple users. In the latter case, the number of users that can simultaneously use the system ranges from a minimum of two to a maximum of twelve, with most projects addressing a number between three and five. A possible reason arises from the studies on group productivity, which indicate higher benefits when working in small groups (Mc Grath, 1984).

In multi-user systems it is relevant to look at how different patterns of group interaction are implemented since specific design choices can generate interactions that are beneficial for learning outcomes (Dillenbourg et al., 2009). Except for the “Climate change” project (Lyons et al., 2012) and “Bar Graph Bouncer” (Charoenying et al., 2012), based on the competition between two users, all other projects are oriented toward collaboration. For example, “Feedyer Alien” distributes different roles among the users: one chooses the correct food and the other brings it to the alien (Johnson-glenberg et al., 2012). Such design generates interdependencies among tasks and consequently facilitates discussion and shared decision-making. Instead projects such as “Archimedes” (Malinverni et al., 2012) and “Wooble” (Kynigos et al., 2010), do not differentiate between users and assign the same role to all of them. However, relevant differences are present: in “Arquimedes” each child acts as an individual entity, whereas in “Wooble” children have to behave as a homogenous entity by coordinating their collective displacement.

Although analyzing the relation between group interaction and learning is not the purpose of this paper, it is necessary to consider the definition of different patterns of collaboration as a fundamental design choice since it can be used both to convey meaning and to facilitate social interactions which are beneficial for learning.

Input technologies and operationalization of user behavior

Social sciences suggest that the notion of body depends on how we look at it (Thomas, 2003). This concept, applied to HCI, requires analyzing how the system conceptualizes the user’s body and his behavior. According to Fogtmann et al. (2008), interaction with a system always imposes constraints on how the body is involved in this interaction. This is due to the choice of different sensing devices which offer diverse understandings of the interacting body.

Two main approaches are used to sense the interacting body: either a “sensor-activation” approach or a “motion tracking” approach. Within the former context different commercial devices are used (e.g., DanceMat, Wiimote,
smart phone’s accelerometer), while motion tracking approaches are based on specifically developed artificial vision systems, Kinect devices or laser scanning. Despite differences in the sensing systems used, most projects are based on tracking body or limbs position in space, with some of them integrating also data on the quality of movements (i.e., speed, flow, time, etc.).

Other projects propose different operationalization of user behavior such as: tracking gestures, the quantity of movement or the collective behavior of multiple users. Examples of this last approach can be found in “Wobble” and “Watergames.” In “Wobble” players control a virtual board through their collective displacement on the floor: the system recognizes the total weight of players and its distribution and operationalizes the interaction in terms of coordinated movement and proximity between users (Kynigos et al., 2010). In “Watergames” the system reacts and activates a fountain only when a group of children creates a closed ring and spins around the fountain (Pares et al., 2005).

These examples, by showing different forms of conceptualizing users, suggest that the definition of the input represents a fundamental starting point. This is because it directly affects the mental models of users on how they can interact with the system and defines possibilities for meaning construction. The definition of the sensing technology does not only represent a technological choice but can feed knowledge construction by operationalizing user behavior in a meaningful way.

**Mapping**

The concept of mapping, defined by Norman (1988) as the relation between “controls and their movements and results in the world”, has been present for many years in HCI. While in its general definition mapping provides a strategy to make an interface easy to use, within FUBILEs a further layer of complexity is added since users do not only have to understand how the systems works but also must understand the target learning goals (Fig. 1).

![Figure 1. Multi-layered mapping. Users need to understand the system (first layer) and mapping should support the understanding of the content (second layer)](image)

A first aspect to analyze is whether mapping and user actions are related to the defined learning goal and how they related. Three main approaches were identified: a “functional” approach, an “identity” approach and a “metaphorical” approach.

The “functional” approach represents the cases in which user actions and mapping do not relate with the content but arise from a usability perspective. Examples can be found in (Gronbaek et al., 2007; M. C. Johnson-glenberg et al., 2010; Kiili et al., 2012; Lim et al., 2011; Lucht & Steffi, 2013), which mainly rely to the application of traditional interaction paradigms of mouse-based interaction (i.e., select, drag, drop). The majority of these projects arise either from a technologically oriented framework or from approaches related to exertion and to the adaptation of existing materials.

On the other hand, both the “identity” and the “metaphorical” approaches are oriented toward establishing a relation
between user actions and content. Based on the Embodied Cognition framework and on a semantic design, these projects focus on designing actions that allow having a physical experience of the concept taking advantage of body knowledge.

The identity approach orients the mapping toward creating tight analogies between user action and content. For instance “Magic Angle,” which aims at the learning of methods to measure angles, requires children to physically enact angles with their arms (Liu et al., 2006). Similarly “Kinematics graphic” (Anastopoulou et al., 2011) maps the arm movement to graph of the relation between velocity, time and distance. Finally “Meteor” (Lindgren & Moshell, 2013) and “Archimedes” (Malinverni et al., 2012) require children to physically perform the behavior of specific virtual objects. From a semantic perspective the identity approach works as a denotative meaning, understood as using the literal meaning of user actions to indicate the concepts.

However, as Holland et al. (2011) pointed out, some concepts cannot be expressed through the literal meaning of body actions. To address this issue the “metaphorical” approach proposes a mapping strategy based on defining metaphors to bond users actions with the content. In these cases user action does not correspond to its literal meaning but works as a symbol for something else; e.g., (Antle et al., 2013, 2008; Carreras & Pares, 2009; Holland et al., 2011; Lyons et al., 2012; Pares et al., 2005). For instance, the project “SpringBoard” uses Johnson’s Embodied Metaphor of balance to map the proprioceptive experience of balancing our own body onto the concept of social justice. The use of a metaphor offers strategies to address abstract concepts that cannot be physically enacted in a literal way; however it is necessary to reflect on whether the multiple layers of meaning of metaphors can facilitate the understanding or add an additional level of complexity in the learning process. For metaphors to be correctly understood, they need to be based on strongly shared socio-cultural references. This is often difficult to guarantee, especially in projects addressed to children.

<table>
<thead>
<tr>
<th>Table 2. Interaction design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
</tr>
<tr>
<td>Physical configuration</td>
</tr>
<tr>
<td>Vertical screen / projection</td>
</tr>
<tr>
<td>Floor projection</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Number of users</td>
</tr>
<tr>
<td>Single users</td>
</tr>
<tr>
<td>Multiple users</td>
</tr>
<tr>
<td>Competitive</td>
</tr>
<tr>
<td>Collaborative</td>
</tr>
<tr>
<td>Input data</td>
</tr>
<tr>
<td>Body/limb position</td>
</tr>
<tr>
<td>Quality of movement</td>
</tr>
<tr>
<td>Quantity of movement</td>
</tr>
<tr>
<td>Gestures</td>
</tr>
<tr>
<td>Collective behavior of multiple users</td>
</tr>
<tr>
<td>Mapping</td>
</tr>
<tr>
<td>Functional</td>
</tr>
<tr>
<td>Identity function</td>
</tr>
<tr>
<td>Metaphorical</td>
</tr>
</tbody>
</table>

Evaluation

Twenty-one papers report empirical studies for evaluating the impact on learning and user experience. Sixteen of them focus on assessing learning outcomes, while five are oriented toward evaluating user experience. In all cases an effort is posed on evaluating the specificities of Full-body Interaction by comparing it with other instructional methods or with traditional interfaces.

Evaluating learning

Papers aimed at assessing the impact on learning are mainly based on the administration of retrospective measures, understood as evaluating learning through the administration of tests after the experience. Two main approaches are used: three studies use only post-task assessment, while eleven use an experimental design based on pre and post test. Both single measure and repeated measures have pros and cons: the latter can increase the risk of repetition biases.
while the former makes it difficult to properly track learning gains.

Only Cress et al. (2010) use data logs of the performance to evaluate learning. In their study on using the DanceMat to foster numerical skills in children, they analyze data logs to assess improvements in accuracy and performance time between the first and second use of the application. Data logging can provide highly reliable quantitative data, however its application for complex learning goals can be challenging.

In terms of assessment instruments nine projects are based on multiple-choice questionnaires, three on interviews and three on mixed methods that combine interviews, questionnaires and discussion. This distribution requires a reflection on the relation between theoretical frameworks and assessment instruments. Most projects rely on constructivism but utilize instruments such as multiple-choice questionnaires. This does not seem coherent since questionnaires are often criticized by constructivism itself due to their deficiencies in properly discriminating the reasoning process behind the answers (Berg & Smith, 1994) and in evaluating understanding at a deeper level (Reeves & Okey, 1996). This contradiction requires further research on evaluation methods for FUBILEs, since only one project explicitly includes the tracking of the learning process and a deep analysis of students’ understanding (Howison et al., 2011).

At the same time retrospective assessment is partially contradictory with the very nature of FUBILEs and embodied learning. Recent studies suggest that implicit bodily-based knowledge may precede the ability of children to properly articulate verbally their understanding (Broaders & Goldin-Meadow, 2010). Such aspect suggests the necessity of considering the role of bodily-based knowledge and overcoming the limits of verbal expression in the assessment of FUBILE.

The assessment of full-body interaction: Comparing interfaces

Ten out of fifteen papers aimed at assessing learning report studies that evaluate the specific potential of Full-body Interaction by comparing it with other interfaces or instructional methods. Such studies are based on experimental designs where users are assigned either to a Full-body Interaction condition or to a control condition.

Relevant differences are present in the control condition. Four studies compare the use of the same application with two different interfaces; e.g., in “Arquimedes”, the same application is used either on a large inflatable slide or on a desktop computer (Malinverni et al.,2012). Two papers compare their system with a different desktop application, i.e., “SpatialEase” compare a FUBILE for learning second language with an already existing desktop application that has the same learning goals (Edge et al., 2013). Finally four papers compare the interactive experience with other instructional methods, such as traditional classroom instructions.

Despite the diversity in their research question, all comparative studies evaluate whether there is a significant difference between two conditions (i.e., whether an interaction modality is more effective than another) but none of them consider the qualitative aspects related to how users can learn differently depending on the interface. Such limitation has been already discussed by Kozma (1994) who suggested deepening the research on the effects of media on learning through the use of more qualitative approaches.

Results

Results of learning outcomes and comparative studies report a heterogeneous panorama. Despite fifteen papers evaluate learning, only fourteen clearly report their results: nine report significant improvements in learning, while five did not find any significant learning gains.

Comparative studies between Full-body Interaction and traditional instructional methods report significant differences in learning gains for the users assigned to the FUBILE condition. Comparative studies between FUBILE and desktop interfaces show no significant difference in five out of six projects. Only Cress et al. (2010) who compared a numerical task on DanceMat and on a tablet computer reported a significant difference in favor of users of the FUBILE.
The broad diversity in learning goals, design strategies and assessment methods makes it difficult to extrapolate consistent conclusions about these findings, however some aspects can be considered to provide guidelines for research and design. The projects that reported significant learning gains addressed learning goals that are clearly operationalized, while most studies that did not report significant gains deal with much more diffuse topics. For instance “Arquimedes”, which reported significant learning gains, focused its learning goal on Archimedes principle (Malinverni et al., 2012); conversely “Wooble”, which did not report any learning gain, addresses goals related to “force, balance, weight, location, direction” (Kynigos et al., 2010). Despite several factors could contribute to the effectiveness of a FUBILE, it could be beneficial to clearly delimitate learning goals to circumscribe the amount of information that we want to communicate to the user. The results of “Wobble” are particularly indicative. After using the system, users were asked to define its educational content and none of them related it with physics. This misunderstanding suggests the necessity of carefully evaluating the communicative and interpretative aspects of design.

Studies on the comparison with traditional methods confirm the literature on the benefits of learning by doing and the role of active experience in grounding concepts. On the other hand, studies on the effect of the interactional difference posit much more challenging questions. As mentioned before, the study proposed by Cress et al. (2010) has two differential elements: (1) it is the only one that uses an evaluation method based on analyzing user performance and (2) it addresses the topic of numerical magnitude, which is deeply related to spatial and bodily cognition.

Even if it is not possible to disentangle these factors, novel research lines for FUBILEs could address the definition of proper learning goals and adequate assessment methods. Relevant shortcomings can found in a dichotomist approach (i.e., “is A more effective than B?”), which does not allow understanding in which aspects Full-body Interaction really differs from other kinds of interaction modalities.

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Significant difference</th>
<th>Non significant difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning gains</td>
<td>(Cress et al., 2010; Edge et al., 2013; Enyedy et al. 2012; Howison et al., 2011; Johnson-glenberg et al., 2010, 2011b; Lucht &amp; Steffi, 2013; Malinverni et al., 2012)</td>
<td>(Carreras &amp; Pares, 2009; Hashagen et al., 2009; Kynigos et al., 2010; Lim et al., 2011)</td>
</tr>
<tr>
<td>Difference between conditions</td>
<td>Different interfaces (Cress et al., 2010)</td>
<td>(Edge et al., 2013; Johnson-glenberg et al., 2010; Lindgren &amp; Moshell, 2013; Malinverni et al., 2012)</td>
</tr>
<tr>
<td>Difference between instructions</td>
<td>Different instructions (Anastopoulou et al., 2011; M. C. Johnson-glenberg et al., 2010, 2011a; Lucht &amp; Steffi, 2013)</td>
<td></td>
</tr>
</tbody>
</table>

**Evaluating user experience**

Seven papers focus on evaluating user experience, while the other combines this evaluation with the assessment of learning. Different aspects are analyzed: the enjoyment of the experience, its impact on the user, the level of motivation, and changes in the attitude toward the learning goals. Methods include interviews, questionnaires and observations. Contrary to assessment of learning, results are highly homogeneous, reporting significant differences in engagement, immersion and motivation in all of the studies aimed at comparing user experience in Full-body Interaction with a traditional interface.

Such findings suggest that FUBILEs can be highly effective in emotionally involving the users and motivating them. Antle et al. (2013) and Lucht et al. (2013) report particularly relevant results. Antle reports a significant difference in awareness, impact and willingness to get involved with social justice between the users assigned to the FUBILE and those assigned to the control condition. Lucht shows that children trained with FUBILE report a significantly higher positive attitude toward English, when compared to those assigned to the traditional classroom condition.
Such findings are consistent with the research on physical activity and user’s attitude (Bianchi-berthouze, 2013) and suggest considering the pedagogical potential of FUBILEs from a perspective that includes the affective aspects of learning. However, longitudinal studies become extremely necessary at this point to be able to compensate for the novelty factor.

**Discussions**

This review reports a panorama of the theoretical frameworks, design approaches and empirical evaluations of FUBILEs, identifying tendencies and shortcomings useful for future research. At the same time the structure of the analysis can serve as a framework for research by providing a “checklist” of critical aspects to consider when designing for FUBILEs (see Table 4).

<table>
<thead>
<tr>
<th>Table 4. Design checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational context</td>
</tr>
<tr>
<td>- Definition of learning goals and learning process: How are the learning goals selected? Is there any theoretical support for their relation with physicality? Do they address needs expressed by stakeholders? How are users supposed to learn?</td>
</tr>
<tr>
<td>- Definition of target users</td>
</tr>
<tr>
<td>- Context: How is the educational experience framed?</td>
</tr>
<tr>
<td>Interaction design</td>
</tr>
<tr>
<td>- Physical interface: Which physical interface will be used? Why? Which affordances does it support?</td>
</tr>
<tr>
<td>- Number of users: How many users will be able to use simultaneously the system? Which patterns of group interaction are implemented? Which emergent social interactions can be provoked?</td>
</tr>
<tr>
<td>- Operationalization of user behavior: which aspects of user behavior are used as inputs by the system? Do these aspects allow mapping with learning goals?</td>
</tr>
<tr>
<td>- Mapping: Do the mapping of user activities relate with learning goals? How?</td>
</tr>
</tbody>
</table>

Three major theoretical frameworks have been identified as bedrocks of research in FUBILEs: (1) a pedagogical framework, based on the importance of learning-by-doing; (2) an Embodied Cognition framework that scientifically grounds the fundamental role of the body in cognitive processes; (3) a physiology-oriented framework that suggests the tight relation between exertion and academic performance. These three theoretical frameworks are reflected in specific design strategies, which include the adaptation of existing materials, the balance between physical workload and cognitive load and the use of elements of the system to communicate meaning. Within this latter approach a number of semiotic resources have been identified: the design of user activity, the role of space, the coupling between physical movement and computational feedback, the notion of effort and the relations between multiple users.

At the same time the analysis of the design strategies showed that most of them are based on a theory-oriented approach, with only few examples including either users or stakeholders in the design process. Such tendency is partially unaligned with the emergent claim of HCI community about involving users at a deeper level. It suggests the need of researching on design methods capable of blending complex theoretical models with a user-centered approach. This need is even more evident in the cases in which the designer’s goal is to communicate abstract concepts.

**Future work**

The analysis of specific design choices allowed formalizing a preliminary structure that can serve as guidelines for researchers and designers. The review of the physical interfaces, number of users, system inputs and mapping showed that these aspects are either designed to accomplish a functional purpose or to facilitate meaning construction. However, even when a design choice is mainly focused toward functionality, it is necessary to be aware of its communicative potential. This implies carefully analyzing the specific affordances, cultural constructs and behavioral effects that it can evoke.
Additional research is needed to properly take into account the broad amount of variables embedded in FUBILEs. Possible research paths can analyze the relation between design choices and the emergence of specific user behavior and affects. Studies in Tangible User Interaction and learning show that even subtle changes in the physical configuration of the environment can have relevant impact on comprehension (Price & Jewitt, 2013), pointing out the need of fine-grain analysis of the relation between design choices, behavior and cognitive processes.

This need is intertwined with the research on adequate evaluation methods. Evaluation methods should be aligned with the underlying theoretical frameworks, while at the same time be capable of offering a deeper understanding on the specific effects of Full-body Interaction on learning outcomes. Relevant possibilities can be found in overcoming the binary approach (i.e., “is A more effective than B?”) to properly analyze which understanding and skills can be fostered through Full-body Interaction. Such approach pushes us towards searching for misconceptions, emerging representations and plotting different levels of learning outcomes. Interesting approaches can be found in combining methods such as Stealth Assessment (Shute, 2005), which evaluates the skills of the learner during the interaction, with the analysis of embodied meaning construction. Finally, research should address the understanding of whether some kind of users can benefit more of a Full-body Interaction experience; e.g., initial low knowledge vs. initial high knowledge or different categories of learners.

References


Quality Indicators for Learning Analytics

Maren Scheffel*, Hendrik Drachsler, Slavi Stoyanov and Marcus Specht

Open Universiteit Nederland, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands // maren.scheffel@ou.nl // hendrik.drachsler@ou.nl // slavi.stoyanov@ou.nl // marcus.specht@ou.nl

*Corresponding author

ABSTRACT

This article proposes a framework of quality indicators for learning analytics that aims to standardise the evaluation of learning analytics tools and to provide a mean to capture evidence for the impact of learning analytics on educational practices in a standardised manner. The criteria of the framework and its quality indicators are based on the results of a Group Concept Mapping study conducted with experts from the field of learning analytics. The outcomes of this study are further extended with findings from a focused literature review.

Keywords

Learning analytics, Quality indicators, Group concept mapping, Framework

Introduction

In the last few years, the research field of learning analytics (LA) has been growing steadily. According to Siemens (2011) LA is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.” Building on ideas from process mining, data processing, information retrieval, technology-enhanced learning, educational data mining, and visualisation LA is a multi-disciplinary research field that now forms its own domain. Several resources and organisations are already dealing with the topic in special journal issues, conferences, workshops as well as courses, summer institutes, and societies (SoLAR, 2014) specifically dedicated to LA. There, the research community has worked on the state of the art in learning analytics, its processes, frameworks, definitions, and challenges (see Clow, 2012; Drachsler & Greller, 2012; Duval, 2011; Elias, 2011; Ferguson, 2012a; Ferguson, 2012b; Greller & Drachsler, 2012; Siemens & Baker, 2012).

Making use of learning analytics can give added value to learners as well as educators. Many university courses today consist of a blended approach between classroom lectures and self-regulated learning activities. LA can help learners to better plan and reflect these activities by becoming aware of their actions and learning processes. According to Endsley (1995, 2000) being aware of one’s own situation is a three level process and a prerequisite for making decisions and effectively performing tasks: the perception of elements in the current situation is followed by the comprehension of the current situation which then leads to the projection of a future status. Once a learner is aware of his situation, he “reflects on the phenomenon before him, and on the prior understandings which have been implicit in his behaviour” (Schön, 1983) to then engage in a process of continuous learning. Reflection can promote insight about something that previously went unnoticed (Bolton, 2010) and lead to a change in learning behaviour. Thus, results of LA can be used to foster awareness and thus reflection (Verpoorten et al., 2011; Verpoorten, 2012; Govaerts et al., 2012) or to give recommendations for further steps in a current learning scenario (Greller & Drachsler, 2012). As Ferguson (2014) explains, LA offers “ways for learners to improve and develop while a course is in progress. These analytics do not focus on things that are easy to measure. Instead, they support the development of crucial skills: reflection, collaboration, linking ideas and writing clearly.” Awareness and reflection support for students are consequently highly important aims of learning analytics. The existence and impact of these aims, however, are hard to measure due to the lack of standards that the student support of LA tools can be measured against.

The same applies to educators. In order to support students within a course, teachers should be aware of what the students are doing, how they are interacting with the course material, where comprehension problems arise (cf. Scheffel et al., 2011; Scheffel et al., 2012). Especially if the number of students in a course is high and the tasks the students are engaged in are not trivial, teachers need assistance for keeping track of the students’ activities, e.g., with the help of activity-based learner-models (Florian et al., 2011). Zinn & Scheuer (2006) conducted a survey among teachers trying to identify requirements for student tracking tools. Among the information deemed mostly important were the students’ overall success rate, the mastery level of concepts, skills, methods and competencies as well as the
most frequently diagnosed mistakes. Such information is also needed for the evaluation of a course, i.e., didactic concept, materials, contents, tools, and tests. Awareness and reflection support for educators are thus also highly important aims of LA. But as with the learner support, standards that define quality indicators for learning analytics tools are missing.

While the added value of LA for learners and educators is clearly recognised (Long & Siemens, 2011), little research has been done so far to compare the findings of empirical LA studies and their tools as having a desirable effect on learning. This article therefore proposes to work toward quality indicators for learning analytics that will help standardise the evaluation of LA tools. It provides a first version of a framework of quality indicators to measure and compare the impact of LA on educational practices.

The quality indicator framework has been developed with experts from the LA domain by using a Group Concept Mapping (GCM) approach. The remaining parts of the article are organised in the following way: First, we will present the GCM methodology and provide some demographic description of the participants. Second, we will present and discuss the empirical findings of the study that reflect the LA community’s view on such quality indicators. Third, we will propose a first version of a framework of quality indicators for learning analytics. Fourth, we will further extend the findings of the GCM study with a focused literature study of related articles. Finally, we will conclude our results and provide some limitations and potential future research directions toward the application of the quality indicators in learning analytics.

**Group concept mapping**

**Method**

One methodology to identify a group’s common understanding of a given issue is Group Concept Mapping. It is a very structured approach that applies quantitative as well as qualitative measures that create a stakeholder-authored visual geography of ideas from a target group, combined with specific analysis and data interpretation methods, to produce maps to guide planning and evaluation efforts on the issues of the group (Kane & Trochim, 2007). Our study makes use of a GCM online tool (Concept Systems Global, 2014) and consists of three steps for the participants: (1) generation of ideas, i.e., quality indicators of learning analytics, (2) sorting of the collected ideas into clusters, and (3) rating of the ideas according to several values, e.g., importance and feasibility. The individual input of the participants is aggregated to reveal shared patterns in the collected data by applying statistical techniques of multidimensional scaling and hierarchical clustering. Visualisations then help to grasp the emerging data structures and to interpret the data. One important aspect of GCM is its bottom-up approach. Instead of presenting a given set of criteria to sort and rate, the community itself generates the ideas that are to be clustered and rated by a group of experts.

**Participants**

The involvement of participants in our GCM study was twofold (see Table 1). The first phase was conducted during the days of the Learning Analytics and Knowledge Conference 2014. Calls for participation were circulated via several channels, e.g., Twitter, project websites, personal contact, email etc., asking people involved and interested in LA to contribute their quality indicators for learning analytics to the brainstorming phase. Participation was accessible via a link and open, i.e., people did not have to register with the GCM tool. In total, 74 people participated in the brainstorming phase.

<table>
<thead>
<tr>
<th></th>
<th>Started</th>
<th>Finished</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brainstorming</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Demographic questions</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>Sorting</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>Rating importance</td>
<td>24</td>
<td>21</td>
</tr>
<tr>
<td>Rating feasibility</td>
<td>22</td>
<td>20</td>
</tr>
</tbody>
</table>
For the second phase, i.e., sorting and rating of the collected quality indicators, we selected 55 experts from the domain of LA (i.e., they had been involved in the domain for several years, had published about learning analytics-related topics, were from the higher education sector and preferably had a PhD degree) and contacted them personally. Table 2 shows a summary of the demographics, the average expert of the study is a researcher at a university with an advanced expertise in LA and has more than six or ten years of work experience.

### Table 2. Answers to demographic questions by participants of phase two

<table>
<thead>
<tr>
<th>Participant Question</th>
<th>Option</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expertise</td>
<td>Novice</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>6</td>
<td>25.00</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>45.83</td>
</tr>
<tr>
<td></td>
<td>Expert</td>
<td>7</td>
<td>29.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>100.00</td>
</tr>
<tr>
<td>Experience</td>
<td>Less than 5 years</td>
<td>8</td>
<td>33.33</td>
</tr>
<tr>
<td></td>
<td>6-10 years</td>
<td>5</td>
<td>20.83</td>
</tr>
<tr>
<td></td>
<td>More than 10 years</td>
<td>11</td>
<td>45.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>100.00</td>
</tr>
<tr>
<td>Involvement</td>
<td>More in research</td>
<td>16</td>
<td>66.67</td>
</tr>
<tr>
<td></td>
<td>More in teaching</td>
<td>1</td>
<td>4.17</td>
</tr>
<tr>
<td></td>
<td>Equal in research and teaching</td>
<td>4</td>
<td>16.67</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>3</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Procedure**

All participants of all three activities were informed about the purpose, the procedure, and the time needed to complete the activities. Participants of the first phase were given a link to access the brainstorming section of the GCM tool and asked to generate ideas by completing the following statement: "One specific quality indicator to evaluate the effects of learning analytics is ...." Participants had ten days to contribute to the brainstorming. During this first phase the 74 participants generated a total of 92 original ideas. Before releasing the list of statements into the second phase, identical statements were unified and too vague ideas, e.g., "Range of flexibility in moving from one point to another in a theoretical discussion," were taken out by us. Also, those statements that contained more than one idea were split, e.g., "students and teachers change their behaviour in some aspects" was split into one statement for teachers and one for students. After the cleaning process, the now 103 statements (the full list is available at http://bit.ly/103QILA) were randomised and pushed into the sorting and rating phase. Participants first sorted the statements according to their view of the statements’ similarity in meaning or theme and to also name the clusters. Dissimilar statements were not to be put into a “miscellaneous” cluster but rather into their own one-statement-cluster in order to ensure statement similarity within the clusters. Then, participants rated all quality indicators on a scale of 1 to 7 according to their importance and feasibility, with 1 being of lowest and 7 being of highest importance/feasibility. Participants had two weeks to complete the sorting and rating activities.

**Results**

**Point map**

The GCM tool offers a number of automated analyses of the collected data: multidimensional scaling and hierarchical clustering for the sorting data and mean, standard deviation and correlation for the rating data. Figure 1 shows a point map of the 103 quality indicators, i.e., the outcome of the multidimensional scaling analysis. The multidimensional scaling analysis assigns a so-called bridging value between 0 and 1 to each statement. Statements with low bridging values have been grouped very close together with other statements around it, e.g., statements 98, 52, 75, 99 on the lower right side of Figure 1 all deal with some form of student motivation and can be considered quite coherent. Statements with higher bridging values can also be grouped together but the surrounding statements are then further apart, e.g., statements 95, 23, 50, 61 about teacher motivation, engagement and feedback.
statements that are close to one another in the map are also close to one another in meaning and have thus been clustered together by the experts.

**Figure 1.** Point map of the 103 quality indicators

From point map to cluster map

In some areas of the map it is quite easy to detect groups by simply looking at the point map. In other areas, however, it is more difficult to decide where group boundaries could be. The hierarchical clustering analysis of the GCM tool offers several solutions to a given point map. We used a cluster replay map, starting at 15 clusters and working down to two (see Figure 2). For each cluster-merging step we carefully looked at the statements of the clusters that were to be combined to check whether a merging made sense. The solution that seemed best to be representing the collected data and the purpose of the study was the one with eight clusters.

**Figure 2.** Replay map showing 15 clusters

After deciding on the number of clusters to work with, meaningful labels needed to be constructed for the clusters. The system automatically suggests a list of labels per cluster. Another way of finding appropriate labels is to look at the bridging values of the statements within a cluster. The lower the bridging values are, the better do those statements define the cluster. A third way to find meaningful cluster labels is to find the overarching theme of a cluster by looking at all statements of a cluster. We combined all three methods to define the labels of the 8-cluster

The GCM tool assigns each cluster with a bridging value. The more coherent a cluster is, the lower is its bridging value which can be attributed to a high agreement rate from the experts about statements within a cluster. The four most coherent clusters are Student awareness (0.11), Data: open access (0.25), Data: privacy and Learning performance (0.31 each). The clusters Teacher awareness (0.41), Learning support (0.45) and Learning outcome (0.46) are all similar in range. The cluster with the by far highest bridging values and thus least coherence is Acceptance & uptake (0.86). In order to get a better grasp of the different clusters, a more detailed description of their characteristic statements is given below.

**Cluster descriptions**

Cluster 1 **Data: open access** contains eleven statements with bridging values ranging from 0.06 to 0.60. Most statements deal with aspects of openness and transparency of the used data as well as the used algorithms, e.g., “that data are open access,” “portability of the collected data.”

Cluster 2 **Data: privacy** is about exactly that: privacy, control of data, and transparency of data access. There are eight statements in the cluster with bridging values ranging from 0.10 to 0.72. Representative statements are “that privacy is ensured,” “if learners can influence which data are provided.”

Cluster 3 about **Acceptance & uptake** contains 13 statements and is very diverse as can be seen from the bridge value range from 0.66 to 1.00. The cluster describes aspects of acceptance of LA and its results by different stakeholders but also the comparability of methods or the context and objectives dependence of LA. An example statement is “that administrators invest in scaling successful tools across their programming.”

Cluster 4 **Learning outcome** is also somewhat diverse with a bridging value range from 0.19 to 0.87. It contains 13 statements that deal with comparability of LA results, teacher motivation, result accuracy and feedback for teachers, e.g., “if teachers are able to gain new insights using the given LA methods,” “that LA results are compared with other (traditional) measures.”

Cluster 5 **Teacher awareness** consists of twelve statements with bridging values from 0.18 to 0.73. Most statements are connected to teachers changing their course material or their teaching behaviour in response to LA results about
theirs students: “that teachers change their behaviour in some aspects,” “that teachers react in a more personalized way to how their students are dealing with learning material.”

Cluster 6 Learning performance is one of the smallest clusters as it consists of only eight statements. The bridging value range is relatively small, i.e., 0.11 to 0.59. Statements in this cluster are about student performance, learning and achievement improvement. Representative statements are “that change in workplace learning is measurable,” “the extent to which the achievement of learning objectives can be demonstrated.”

Cluster 7 Learning support is a very stable but also rather large cluster with 18 statements. Its bridging values range from 0.14 to 0.76. Statements in this cluster are often formulated generally and deal with support for teachers as well as for students, e.g., “an early detection of students at risk,” “the ability to explain what could help to further improve,” “that students regularly utilize the tools provided.”

Cluster 8 about Student awareness is the largest and most coherent cluster. It contains 20 statements and its bridging value range is from 0.00 to 0.43. The cluster is also very stable and consistent. All statements are related to students, their achievement, success, self-regulation, awareness, learning behaviour and motivation, e.g., “that students become more self-regulated in their learning processes,” “that students are more aware of their learning progress.”

Rating maps

Once the cluster map is settled upon, the experts’ ratings of the quality indicators can be included in the calculation as well. Two aspects were given to the experts to be rated on a scale from one to seven (one for a low, seven for a high rating): importance and feasibility. While the former refers to the priority or importance of a quality indicator in relation to the evaluation of effects of LA, the latter indicates the perceived ease of applicability of a quality indicator. The GCM tool automatically applies the experts’ ratings to the cluster map, indicating the importance or feasibility by layering the clusters. The system always divides the ratings into five layers based on the average ratings provided by the participants for the rating maps. The anchors for the map legend are based on the high and low average ratings across all of the participants. One layer indicates a low rating, whereas five layers indicate a high rating of the respective aspect.

Figure 4 shows the rating map according to the importance aspect. Clusters Data: privacy, Learning outcome, Learning support, and Student awareness each received very high importance ratings as they all have five layers. Teacher awareness has three layers, while Data: open access and Learning performance have two layers each and Acceptance & uptake, i.e., the least coherent cluster, has only one layer.
Looking at the feasibility-rating map (see Figure 5) one can see a change in the rating behaviour of the experts. Although the Data: privacy cluster also gets five layers and is thus deemed highly feasible by the experts, the other three very important clusters have been rated less feasible: Learning outcome and Learning support only receive an intermediate level of feasibility with three layers each. Student awareness, a highly important cluster, receives a low feasibility rating with two layers only. Teacher awareness also drops down to two layers. The cluster dealing with Acceptance & uptake was seen as neither important nor feasible by the experts. The only cluster that receives more layers in the feasibility-rating map is Data: open access and is thus deemed more feasible than it is important.

A ladder graph (see Figure 6) offers a form of visualisation that is well suited to compare the clusters’ ratings according to importance and feasibility. The rating values are based on a cluster’s average rating. A Pearson product-moment correlation coefficient \( r = 0.65 \) indicates a strong positive relationship between the two aspects of importance and feasibility. For both ratings, the Data: privacy cluster receives the highest values while Acceptance & uptake receives the lowest. As was already observable from the rating maps, the three clusters about Learning outcome, Learning support and Student awareness have all been rated as very important but as much less feasible.

A third visualisation the GCM tool offers for the rating aspects are so-called go-zones, i.e., bivariate graphs that allow to explore the statements in relation to their ratings more deeply. A Go-zone graph maps each statement onto a space between x- and y-axis based on the mean values of the two rating aspects of importance and feasibility. Go-
zone graphs can be created for all statements together or for individual clusters. Figure 7 shows the go-zone graph for all 103 statements. Go-zone graphs are very supportive for the selection of suitable quality indicators for the framework as they highlight those statements with a good balance of importance and feasibility. When deciding on the quality indicators for the framework it can also be sensible to choose statements from the only feasible or only important quadrant if they are close enough to the upper right quadrant and support the criterion.

Figure 7. Go-zone graph of all 103 statements

Constructing a framework of quality Indicators

Discussion of the group concept mapping outcomes

Looking at the clusters in Figure 3, their coherence is also observable visually. One can see that the four most coherent clusters (Data: open access, Data: privacy, Learning performance and Student awareness), i.e., the ones with smaller bridging values, are the smaller ones in relation to area size and that the three least coherent clusters (Acceptance & uptake, Learning outcome and Learning support), i.e., the ones with higher bridging values, are much larger in area size. The two most stable clusters are the ones about Learning support and Student awareness, i.e., they both remained stable until the five-cluster solution while the others merged. This implies a fairly high agreement between the experts’ sorting and the system’s multidimensional scaling and hierarchical clustering. We therefore take cluster coherence and stability to be a first indication of relevance when trying to find quality indicators for learning analytics.

Also very interesting conclusions can be drawn when comparing the two rating maps (see Figure 4 and Figure 5) with one another. The Acceptance & uptake cluster received low ratings for importance as well as for feasibility. The experts’ low rating is also supported by the cluster’s coherence. With an average bridging value of 0.86 and individual statement bridging values spanning from 0.66 up to 1.00, the cluster contains a rather diverse collection of statements. When trying to create quality indicators to evaluate effects of learning analytics we therefore focused on all other clusters first in order to find suitable criteria before taking this cluster into account as the indicators it contains are too incoherent, too vague, unimportant and unfeasible as a group.

This leaves us with a slightly different but nonetheless very interesting cluster landscape: The two clusters in the North (1, 2) both deal with data, access, methods, algorithms, transparency and privacy, i.e., with technical issues, while the clusters in the South (5, 6, 8) deal with awareness, reflection, performance and behavioural change of students and teachers, i.e., with human issues. The “technical North” (Data: open access and Data: privacy) and the “human South” (Teacher awareness, Learning performance and Student awareness) are bridged by a wide layer of learning-related clusters (Learning outcome and Learning support). Apart from the North-South view, one can also look at the map with an East-West perspective: The three Eastern clusters (Data: privacy, Learning support and Student awareness) are more concerned with issues during the learning process while the Western clusters (Data:
open access, Learning outcomes, Teacher awareness and Learning performance) are slightly more concerned with issues of learning output and results. This division is of course not to be seen strictly, but these groupings clearly show a thematic tendency. As for the construction of the framework, we conclude that the aspects of technology, stakeholders (humans), learning processes and learning outcomes should all be reflected in the criteria.

Taking the two rating aspects importance and feasibility into account, we get two different versions of the landscape described above. The importance map on the left side of Figure 8 clearly shows that the learning-related middle layer, i.e., the clusters about Learning outcome and Learning support within the dashed line, is deemed highly important by the LA experts. But all Eastern clusters, i.e., the ones about Data: privacy, Learning support and Student awareness within the dotted line, also receive five layers of importance. Generally one can thus say that the focus of importance is on the learning process-related clusters. For the feasibility map on the right side of Figure 8 the landscape shifts. Now there is a clear North-South divide: The technically-oriented clusters in the North (dotted circle) are deemed most feasible by the experts, followed by the learning-related layer in the middle (short dashed circle) and concluded by the human-related clusters in the South (long dashed circle). This again supports the construction of the framework’s criteria according to the data-learning support & process-stakeholder view.

Looking at the ladder graph in Figure 6 allows a closer look at the differences in average ratings for the different clusters. Especially the drop in feasibility compared to importance for a number of clusters is quite obvious. The most striking drop is that of the Student awareness cluster. The experts think student awareness to be quite an important aspect to take into consideration when evaluating effects of LA theoretically but deem it difficult to apply in real world settings. This can be explained with the fact that many teachers, and thus very likely also the LA experts involved in this study, do not think students to be capable enough of judging their own learning processes and progresses as it has been identified by Drachsler & Greller (2012).

When deciding upon the criteria of the framework it is important to find a good trade-off between the importance ratings and the feasibility ratings. Due to the high importance of the clusters about Data: privacy, Learning outcome, Learning support and Student awareness it seems to be sensible to use them as a basis for the criteria of the framework. The feasibility ratings of the clusters can then be used to associate the remaining clusters with these four criteria: The Data: privacy cluster is by far the most feasible one, followed by the Data: open access cluster. The two can thus be combined into one data criterion. The next two clusters on the feasibility scale are Learning support and Learning outcome, two of the criteria candidates that stay on their own due to their high importance rating. As the latter cluster is followed by the Learning performance cluster and as they both deal with learning results and effects, it makes sense to construct a combined criterion from them. The next cluster on the feasibility scale is Student awareness, closely followed by Teacher awareness. Both of them are “human clusters” and concerned with awareness, reflection and behavioural change. They can therefore be combined into one criterion even though they address different stakeholders.
Outline of the framework

From the results of the GCM study we can identify four topic areas that can be turned into criteria for the framework: the first deals with anything related to data, algorithms, transparency and privacy. It is based on the clusters Data: privacy and Data: open access. For the sake of simplicity the criterion is called Data Aspects. It contains the quality indicators Transparency, Data Standards, Data Ownership, and Privacy. The second topic area concerns support for students and teachers during the learning process, i.e., while using LA tools. It is entirely based on the Learning Support cluster and also takes over this name. The quality indicators of this criterion are Perceived Usefulness, Recommendation, Activity Classification, and Detection of Students at Risk. The third topic area deals with the results at the end of the learning process, i.e. any issues of output, consequence, performance, outcome etc. In this case, however, it is not primarily to be seen in relation to individual student performance, e.g., their grades, but refers to the LA tools’ results and outcomes. It is comprised of the two clusters Learning outcome and Learning performance. The criterion is named Learning Measures and Output and contains the quality indicators Comparability, Effectiveness, Efficiency, and Helpfulness. The fourth topic area contains the quality indicators Awareness, Reflection, Motivation, and Behavioural Change of students and educators during the learning processes, i.e., it is about the educational aims identified at the beginning of this article. This criterion is called Objectives.

Most statements related to stakeholders are about learners and teachers and hardly about institutions. This is partly due to the fact that we did not take the cluster Acceptance & uptake that contains some statements about this into account right away. It is also due to such statements being spread over all clusters. As we consider indicators of organisational issues to be an important aspect when considering the evaluation of LA tools (cf. Arnold et al., 2014a), we decided to add a fifth criterion to the framework: Organisational Aspects containing the quality indicators Availability, Implementation, Training of Educational Stakeholders, and Organisational Change.

The quality indicators are based on a review of the statements in the go-zone graphs of each cluster. In most cases these indicators can be found in the upper right quadrant of the go-zone graphs. In some cases statements from the only feasible or only important quadrant were chosen as well if they are close to the important and feasible quadrant. The statements chosen for each criterion are then combined and turned into slightly shorter, more general statements that clearly represent a quality indicator for a given criterion. Figure 9 shows a first outline of the five criteria with their four quality indicators.

![Figure 9. First outline of the framework of quality indicators for learning analytics](image_url)

Literature supporting the criteria

In this section we present a focused literature review to further extend the GCM study with the latest insights from the LA community. It is structured according to the different layers of the framework. Although the three criteria
Objectives, Learning Support, and Learning Measures and Output are clearly separable from one another in regards to their quality indicators and purposes, it is often difficult to exclusively attribute findings from the literature to one of the criteria only. Literature concerning these three layers has therefore been combined into one section.

Objectives, Learning support, and learning measures and output

Some works on awareness (e.g., Endsley, 1995, 2000; Charlton, 2000) and reflection (e.g., Schön, 1983; Bolton, 2010) have already been mentioned in the introduction of this article. They all deal with these educational and pedagogical concepts in general and are not directly attached to the domain of LA. Their findings, however, matter to this domain. McAlpine & Weston’s (2000) work also deals with reflection as a general concept in educational settings. They argue that “reflection is not an end in itself, but a mechanism for improving teaching and hence maximizing learning.”

Studies in the related domain of technology-enhanced learning reveal several aspects that can be used for outcome measurement of recommender systems (Drachsler et al., 2009) but could also be used for the analysis of other educational technologies and LA tools. The first measurement category is a technical one with the parameters of accuracy, coverage and performance. The second measurement category covers educational aspects and involves the parameters of effectiveness, efficiency, satisfaction and drop-out rate. Social network measures form the third category with parameters of variety, centrality, closeness and cohesion.

Clow (2012) points out that “learning analytics should generate metrics that relate to what is valued in the learning process. If the final assessment rewards undesired behaviour, improving the control system to more effectively optimise the results will make the learning worse.” Clow therefore identifies three strategies by which the effectiveness of LA can be improved: (1) enhancement of the speed of response, e.g., real-time feedback rather than summarising feedback, (2) enhancement of the scale of response, e.g., feedback to more than one stakeholder, and (3) improvement of the quality of an intervention, e.g., testing of the intervention or participation of more stakeholders.

Course Signals is an early intervention solution for collegiate faculty (Arnold & Pistilli, 2012) and serves as an example tool of implemented LA. With this tool, teachers can provide feedback to students about their performance and predicted progress. The feedback is comprised of a personalised email and a progress visualisation, i.e., a traffic light signal. Courses that used the tool showed a strong increase in positive grades and at the same time a decrease in negative grades and withdrawals. Both with teachers and with students, Course Signals received positive overall experiences although teachers approached it with more caution than students.

This caution can be set in relation to the findings of a survey among teachers and researchers of LA. The study revealed that “trust in learning analytics algorithms is not well developed” yet (Drachsler & Greller, 2012). Many educators hesitate to take the calculations of algorithms about learning and educational effects as valid while at the same time they hope to gain new insights from those analytics results. The study also showed that for many participants the application of LA cannot provide a more objective assessment than they could do on their own and that a proper assessment of a learner’s state of knowledge is not possible.

A combination of LA and action research to support teachers in educational settings is presented by Dyckhoff et al. (2013). They describe possible effects of learning analytics on teaching and investigate how this could be evaluated. For them, LA tools should be useful to achieve the set goals in a given context. Their findings show that in many cases LA tools do not yet answer all of the questions that teachers have in regard to their educational setting. This especially concerns qualitative analysis as well as data correlation from different source. Quantitative results, however, are often easily available. Among others, the authors relate these shortcomings to an insufficient involvement of teachers in the design process of LA tools and the lack of appropriate, diverse data sources, e.g., student profile data, mobile data. They conclude that “there is a necessity for creation of evaluation tools to measure the impact and effects of LA on the learning process” and for mechanisms to support and reassure awareness and reflection, as well as to improve teaching processes.
An example of work dealing with the design of pedagogical interventions is that of Wise (2014). The author presents four principles of pedagogical learning analytics intervention design: (1) integration, (2) agency, (3) reference frame, and (4) dialogue. Teachers and course developers can build upon these principles in order to support students in productively making use of LA. Wise also describes three core processes that students should be engaged in: (1) grounding, (2) goal-setting, and (3) reflection. The principles together with the core processes form a model of pedagogical LA intervention design.

Data aspects

One important aspect when dealing with data-related criteria is the availability of data sets and data standards. While a few years ago, open access to data sets was hardly constituted (Drachsler et al., 2010), the last few years have shown an immense rise in the availability of and open access to data sets for the technology-enhanced learning, LA and educational data mining domains. Verbert et al. (2012) provide an overview of existing datasets and analyse them along the dimensions of their framework for the analysis of educational datasets. There are three dimensions: (1) dataset properties, (2) data properties, and (3) learning analytics objectives. Several initiatives are now offering access to educational data sets such as the LinkedUp Project (2014) with its LinkedUp Dataset and LinkedUp Data Challenge, the LAK Dataset (Taibi & Dietze, 2013), the DataHub (2014) and the PSLC DataShop (Koedinger et al., 2010).

A number of legal, risk and ethical issues that should be taken into account when implementing LA at educational institutions in the United Kingdom is presented by Kay et al. (2012). They describe that these institutions have to find a balanced way to assure educational benefits, that they are under as much competitive pressure as organisations in the consumer world and that they need to satisfy the expectations of the now arising born digital generations of learners. The authors suggest four principles that provide good practice when tackling the above-mentioned conflicts: (1) clarity, (2) comfort and care, (3) choice and consent, and (4) consequence and complaint.

Willis et al. (2013) apply The Potter Box, i.e., an ethical model in business communications, to LA. They conclude that institutions will have to “balance faculty expectations, various federal privacy laws, and the institution's own philosophy of student development. It is therefore critical that institutions understand the dynamic nature of academic success and retention, provide an environment for open dialogue, and develop practices and policies to address these issues.”

During the EDUCAUSE IPAS Summit in 2013 participants were asked to discuss issues associated with managing risk in student success systems and to identify opportunities for the development of such systems (EDUCAUSE IPAS Summit Report, 2014). More specifically, the discussions focused on three aspects: (1) the identification of internal and external drivers that encourage the implementation of LA, (2) the identification of institutional risks, documentation of effective practices and review of existing and new solutions, and (3) the development of strategies that already take risk issues into account during the design of LA processes. The authors conclude that existing and new data sources have to be integrated in a better way and that educational institutions should know exactly which data they collect for what purpose and who has access to that data. Institutions should also address the movement of students and their data between institutions and should not misuse the collected data to predetermine a student's success.

An analysis of how privacy and ethical issues specific to the context of learning analytics and its related research as well as guidelines about how to comply with common privacy principles are presented by Pardo & Siemens (2014). These principles are conceived from the review of LA proposals, government frameworks and regulatory directives and allow educational institutions to assess their current level of compliance in order to then possibly improve their privacy-related matters. The principles are: (1) transparency, (2) student control over data, (3) right of access / security, and (4) accountability and assessment.

Also relevant for the Data Features criterion is the methodology based on value-sensitive design that incorporates ethical and legal considerations and requirements throughout the research and development cycle of technology as Friedman (1997) explains. Value-sensitive design is the idea that ethical analysis and reflection needs to take place when and where it can make a difference for the design and governance of technology: starting early on in the design and development process, and close to where the technology is being shaped and designed. Ethical considerations
concern first of all the privacy of individuals taking part in the system. A high degree of configurability, the provision of meaningful default options that relate to a privacy-by-default approach, combined with informative explanations given to users are some of the ingredients that will allow the achievement of the notion of informed consent (Van den Hoven, 2008).

Organisational aspects

In the 21st century, more and more higher education organisations apply LA to optimise student success. According to Norris & Baer (2013) such intelligent investments from the organisations have a strong and justifiable return on investment: the implementation of enhanced analytics is to be seen as critical for student success on the one hand and achieving institutional effectiveness on the other as without it, organisations cannot meet the current gold standard for institutional leadership. Norris & Baer conducted a survey among institutional practitioners and vendors about the building capacity in analytics to improve student success and how they determine the state of practice and gaps between needs and solutions. They interviewed 40 leading institutions from the American higher education sector as well as 20 technology vendors and came up with a framework for optimising student success through analytics that contains seven elements: (1) manage the student pipeline, (2) eliminate impediments to retention and student success, (3) utilise dynamic, predictive analytics to respond to at-risk behaviour, (4) evolve learner relationship management systems, (5) create personalised learning environments/learning analytics, (6) engage in large-scale data mining, and (7) extend student success to include learning, workforce, and life success.

In their discussion paper, Siemens et al. (2013) present a national LA strategy to the Australian Government after undertaking a four step process: First, they evaluated the benefits of analytics in other sectors than education, then had a closer look at the data collection policies on provincial, territory, state and national level, followed by a review of universities around the world that are already developing analytics strategies, and finally, they inspected the role corporate partners can play in helping universities achieve analytics competence. Their five final suggestions are: (1) Australian higher education leaders should coordinate a high level learning analytics task force with a variety of stakeholders, (2) existing national data and analytics strategies should be leveraged, (3) guidelines for privacy and ethics should be established, (4) a coordinated leadership program should be set up, and (5) open and shared analytics curricula should be developed with the learning analytics community. Although their paper focuses on the LA situation in Australia, the findings can be applied to other countries as well.

Arnold et al. (2014a) tackle the readiness of institutions to implement LA. Instead of only looking at the maturity of an institution’s already implemented LA solution, the authors try to investigate how institutions that do not apply any analytics yet can become mature to do so. Their Learning Analytics Readiness Instrument (LARI) survey was conducted at nine higher education institutions and focuses on five readiness components for LA implementations: (1) ability, (2) data, (3) culture and process, (4) governance and infrastructure, and (5) overall readiness perception.

With the help of LA educational institutions are able to tune or correct the inner workings of their programs. Méndez et al. (2014) present five techniques that allow institutions to gain such insights: (1) difficulty estimation, (2) dependence estimation, (3) curriculum coherence, (4) dropout and enrolling paths, and (5) a load/performance graph. For their example analysis the authors used data from 2543 undergraduate computer science students at the ESPOL University in Ecuador spanning from 1978 until 2012. With their large study the authors want to show how simple analytics can be used to re-design whole program curricula.

Finally, in their panel discussion at LAK 2014, Arnold et al. (2014b) argue that “in order to truly transform education, learning analytics must scale and become institutionalized at multiple levels throughout an educational system.” During the discussion, panel participants focused on five areas related to the adoption of LA: (1) technology infrastructure, analytics tools and applications, (2) policies, processes, practices and workflows, (3) values and skills, (4) culture and behaviour, and (5) leadership. From the discussed case studies the authors conclude that institutions have to put effort and intention into planning the implementation and adoption of LA. They suggest using existing research and theory as a foundation when beginning to build new theories and research about system level thinking.
Conclusions

This article proposes a first outline of a five-dimensional framework of quality indicators for learning analytics to help standardize the evaluation of LA tools. The work was motivated by the lack of evaluation standards that define quality indicators of LA tools. After introducing the objectives of LA, we presented a GCM study with experts from the LA domain to identify a list of quality indicators. With the help of a number of analysis steps within the GCM tool we first created a point map of the statements that we then turned into a cluster map including cluster labels. The experts’ ratings on importance and feasibility of the statements allowed us to further narrow down the list of possible quality criteria as well as indicators. After taking the rating maps, the ladder graph and the go-zone graphs into account, we were able to propose a first outline of the framework (see Figure 9) with the following five criteria and quality indicators: Objectives (Awareness, Reflection, Motivation, Behavioural Change), Learning Support (Perceived Usefulness, Recommendation, Activity Classification, Detection of Students at Risk), Learning Measures and Output (Comparability, Effectiveness, Efficiency, Helpfulness), Data Aspects (Transparency, Data Standards, Data Ownership, Privacy) and Organisational Aspects (Availability, Implementation, Training of Educational Stakeholders, Organisational Change). In order to extend the found criteria we conducted a focused literature review to show their usage within the community so far.

Limitations of our current approach are related to the participants of our GCM study: Most participants work at a university and are more research- than practice-oriented. It would be interesting to see whether and how the framework and its quality indicators would change if (high) school teachers and/or more practice-oriented university staff were involved in the process.

For our future research we first aim to further extend the literature study to complement the quality indicator framework by analysing the most recent empirical studies on LA according to their evaluation criteria and quality indicators. Ultimately, we aim to transfer the findings into a concrete evaluation instrument that also includes methods for LA stakeholders to test the indicators. Initial applications of the evaluation instrument will be done in higher education institutions first but we will also try to collect experiences in the K12 and commercial sector. A good mean for this approach is the Learning Analytics Community Exchange project (LACE, 2014) that focuses on the exchange of best practises within these sectors.

References


Augmented Reality Trends in Education: A Systematic Review of Research and Applications

Jorge Bacca¹*, Silvia Baldiris¹, Ramon Fabregat¹, Sabine Graf² and Kinshuk²

¹University of Girona, Institute of Informatics and Applications, Av. Lluis Santalo s/n, 17071 Girona, Spain // ²Athabasca University, School of Computing and Information Systems, 1200, 10011-109 Street, Edmonton, AB T5J-3S8, Canada // jorge.bacca@udg.edu // ramon.fabregat@udg.edu // baldiris@eia.udg.edu // sabineg@athabascau.ca // kinshuk@athabascau.ca

*Corresponding author

ABSTRACT
In recent years, there has been an increasing interest in applying Augmented Reality (AR) to create unique educational settings. So far, however, there is a lack of review studies with focus on investigating factors such as: the uses, advantages, limitations, effectiveness, challenges and features of augmented reality in educational settings. Personalization for promoting an inclusive learning using AR is also a growing area of interest. This paper reports a systematic review of literature on augmented reality in educational settings considering the factors mentioned before. In total, 32 studies published between 2003 and 2013 in 6 indexed journals were analyzed. The main findings from this review provide the current state of the art on research in AR in education. Furthermore, the paper discusses trends and the vision towards the future and opportunities for further research in augmented reality for educational settings.

Keywords
Augmented reality, Systematic review, Trends of AR, Personalization, Inclusive learning in augmented reality

Introduction and definitions

In recent years, technology-enhanced learning (TEL) research has increasingly focused on emergent technologies such as augmented reality, ubiquitous learning (u-leaarning), mobile learning (m-learning), serious games and learning analytics for improving the satisfaction and experiences of the users in enriched multimodal learning environments (Johnson, Adams Becker, Estrada, & Freeman, 2014). These researches take advantage of technological innovations in hardware and software for mobile devices and their increasing popularity among people as well as the significant development of user modeling and personalization processes which place the student at the center of the learning process. In particular, augmented reality (AR) research has matured to a level that its applications can now be found in both mobile and non-mobile devices. Research on AR has also demonstrated its extreme usefulness for increasing the student motivation in the learning process (Liu & Chu, 2010; Di Serio et al., 2013; Jara et al., 2011; Bujak et al., 2013; Chang et al., 2014).

An AR system allows for combining or “supplementing” real world objects with virtual objects or superimposed information. As a result virtual objects seem to coexist in the same space with the real world (Azuma et al., 2001). However, AR is not restricted only to the sense of sight; it can be applied to all senses such as hearing, touch and smell (Azuma et al., 2001). AR allows for combining virtual content with the real world seamlessly (Azuma, Billinghurst, & Klinker, 2011). This differs from the notion of a Virtual Environment (VE) where the user is completely immersed inside a synthetic environment. In this sense, “AR supplements reality, rather than completely replacing it” (Azuma, 1997). The Reality-Virtuality continuum (Milgram, Takemura, Utsumi, & Kishino, 1995) clearly shows the relation between a real environment, AR and a virtual environment.

As an example of the current AR applications in education, Ibáñez, Di Serio, Villarán, & Delgado Kloos (2014) created an AR application for teaching the basic concepts of electromagnetism. In this application students can explore the effects of a magnetic field. For that purpose, the components used in the experiment (cable, magnets, battery, etc.) can be recognized using the camera of a mobile device like a tablet. As a result students can see superimposed information such as the electromagnetic forces or the circuit behavior using the tablet. The results of this research show that AR improved academic achievement and provided instant feedback.

Some researchers have proposed different definitions of AR. For example, El Sayed, Zayed, & Sharawy (2011) assert that AR enables the addition of missing information in real life by adding virtual objects to real scenes. Supporting this definition, Chen & Tsai (2012) point out that AR allows for interaction with 2D or 3D virtual objects.
integrated in a real-world environment. Cuendet, Bonnard, Do-Lenh, & Dillenbourg, (2013) argue that “AR refers to technologies that project digital materials onto real world objects.” These definitions are based on one of the features of AR that is the possibility of superimposing virtual information to real objects. On the other hand, a broader perspective has been adopted in the study of Wojciechowski & Cellary (2013). They define AR as an extension of virtual reality with some advantages over virtual reality.

**Current state of AR applications in education**

A considerable amount of literature has been published in AR’s application in educational contexts for a wide variety of learning domains. However, the state of current research in AR for education is still in its infancy (Wu, Lee, Chang, & Liang, 2013; Cheng & Tsai, 2012). According to Wu et al., (2013a) and Cheng & Tsai (2012) the research in this field should continue and should be addressed to discover the affordances and characteristics of AR in education that differentiate this technology from others. Deepening this analysis will allow for discovering the unique value of the learning environments based on AR. According to Chen & Tsai (2012) the potential of AR in educational applications is just now being explored. Dunleavy, Dede, & Mitchell (2009) point out that “we are only beginning to understand effective instructional designs for this emerging technology.”

Table 1 summarizes some review studies available in the literature on the topics related to AR in education.

<table>
<thead>
<tr>
<th>Study</th>
<th>Analysis dimension</th>
<th>Studies reviewed</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Martin et al., 2011)</td>
<td>The review considered the evolution of technology trends in education from 2004 to 2014 through a bibliometric analysis of the Horizon Reports on the topic of AR as well as on other topics of technology enhanced learning.</td>
<td>10</td>
<td>The number of articles about AR is increasing but according to the analysis this technology is in their initial stage in education. In the study the evolution of AR to mobile augmented reality is considered a successful metatrend.</td>
</tr>
<tr>
<td>(Radu, 2012; Radu, 2014)</td>
<td>Review of studies that compare student learning in AR versus non-AR applications.</td>
<td>32, 26</td>
<td>The findings on the positive impact are: Increased content understanding, Learning spatial structures, language associations, long-term memory retention, Improved collaboration and motivation. The findings on the negative impact are: attention tunneling, Usability difficulties, ineffective classroom integration, learner differences.</td>
</tr>
<tr>
<td>(Santos et al., 2014)</td>
<td>The review considered papers published in IEEE Xplore. Authors applied a meta-analysis and a qualitative analysis in the dimensions of display metaphors, content creation and evaluation techniques.</td>
<td>87</td>
<td>Authors conclude that there are three main affordances of AR: real world annotation, contextual visualization and vision-haptic visualization. Also authors state that the three affordances are supported by existing theories like: multimedia learning theory, experiential learning and animate vision theory.</td>
</tr>
</tbody>
</table>

A large and growing body of literature has reported factors such as: uses, purposes, advantages, limitations, effectiveness and affordances of AR when they are applied in different learning domains. However, there is gap in the literature with respect to systematic literature reviews looking at these factors of AR in educational settings. Taking into account this, the aim of this systematic literature review is to present the current status of research in AR in education. The study considers categories for analyzing the current state and tendencies of AR such as the uses of AR in educational settings as well as its advantages, limitations, effectiveness; the availability of adaptation and
personalization processes in AR educational applications as well as the use of AR for addressing the special needs of students in diverse contexts. The analysis of the different categories allows suggesting trends, challenges, affordances, the opportunities for further research and a general vision towards the future.

The rest of the paper is organized in five sections. First section describes the research questions addressed in this systematic review. Second section describes the methodological design of the study. Third section presents the results jointly with the discussion of the findings. Fourth section follows with a discussion on the trends and the vision toward the future. Finally, fifth section remarks some conclusions.

**Research questions**

There is a large volume of published studies that report advantages, limitations, effectiveness challenges, etc. of AR in education. However, since AR is an emergent technology, it is important to get an overview of the advances and real impact of its use in educational settings, describing how AR has been used for generate more student-center learning scenarios. Within this context the research questions addressed by this study are:

- What are the uses, purposes, advantages, limitations, effectiveness and affordances of augmented reality in educational settings?
- Have the inclusion of user modeling and adaptive processes been considered in augmented reality applications?
- How has augmented reality addressed the special needs of access and people preferences in educational settings?
- What are the evaluation methods considered for augmented reality applications in educational scenarios?

**Method**

For this review, we considered the guidelines proposed by (Kitchenham, 2004) and adapted to this literature review:

**Planning:**
- Selection of Journals
- Definition of inclusion and exclusion criteria of studies
- Definition categories for the analysis

**Conduct the review:**
- Study selection
- Data extraction (Content analysis method was applied)
- Data synthesis
- Data coding

**Reporting the review:** This step includes the analysis of results, discussion of findings, trends and conclusions of the review.

Regarding step 3 (Reporting the review), we followed the recommendations of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Moher, Liberati, Tetzlaff, & Altman, 2009). The PRISMA statement is the international and accepted updated version of the QUORUM (Quality of Reporting of Meta-analysis) statement. In the following sub-sections we depict the most important steps followed according to the methodology.

**Step 1a: Selection of journals**

The aim of this step has been to choose the most relevant journals for the systematic review in a consistent way. To keep the process methodologically strong and scientifically consistent, a method has been defined in this research for selecting journals.
The Google Scholar h5-index for the category “Educational technology” was used as the starting point. This starting point was selected since this category is more specific than “Education and educational research” category from the Journal Citation Report Social Science Citation Index (JCR SSCI). In the later, most of the journals about educational technology are indexed jointly with journals about educational research in general.

We chose the top 5 journals from “Educational Technology” category from Google Scholar h5-index and we named this list “GS list.” In order to validate our initial “GS list,” we performed an iterative double check process using the JCR SSCI tool in order to consider the impact factor of each journal and its “relatedness” with others. The “relatedness” or most related journals is defined in the JCR taking into account the cited and citing relationship of the journals and is based on the number of citations from one journal to the other and the total number of articles. The iterative double check process was performed as follows: For each journal in the GS list, we searched the most related journals to that one using the option “Related Journals” in the JCR SSCI web application (Journal Citation Reports - ISI Web of Knowledge, 2012). As a result, we obtained one list of related journals for each journal in the GS list. In this way, we obtained five lists of related journals which were named as RJ-GS1, RJ-GS2, RJ-GS3, RJ-GS4 and RJ-GS5, where RJ stands for “related journal” and GS# stands for the corresponding journal from the GS list.

We then independently sorted each of the lists RJ-GS1 to RJ-GS5 taking into account the impact factor. This process is somewhat similar (by analogy) to a precipitation process (Gooch, 2007) were the journals with major impact factor will “float” in each list. As a result, we obtained 5 independent lists of journals ordered by impact factor. Despite the fact that lists were organized by impact factor, we had some similar journals in each list but at different positions. For example, the British Journal of Educational Technology was at position 7 in the list RJ-GS1 but at position 4 in RJ-GS2. In the remaining lists (RJ-GS3, RJ-GS4 and RJ-GS5) the journal was also in different positions. In order to overcome this situation we combined all the elements of the lists (from RJ-GS1 to RJ-GS5) by pondering the position occupied by each journal through the five lists. As a result, the definite list of journals ordered according to its position was obtained. This list was named FL-JCR-SSCI list.

We then analyzed each journal from the FL-JCR-SSCI list and discarded journals that did not cover topics about educational technology. This analysis was based on the “subject categories” reported for each journal in the JCR SSCI web application. If necessary we analyzed the aim and scope of each journal to see if the journal could be considered. As a result of this process, we had a new list of journals named ET-FL-JCR-SSCI. Where “ET” stands for Educational Technology. This list contains only journals that cover the topic of Educational Technology ordered by impact factor. Table 2 shows the first 5 journals of ET-FL-JCR-SSCI list that corresponds to the journals selected for this review. We have to point out that this method allowed us to find the most important journals in educational technology through a double check process considering impact factor and “relatedness” in the JCR SSCI.

| Table 2. List of the first 5 journals of “ET-FL-JCR-SSCI list” |
|------------------|------------------|
| Journal title | Impact factor (JCR SSCI 2012) |
| Computers and Education | 2,775 |
| Internet and Higher Education | 2,013 |
| British Journal Of Educational Technology | 1,313 |
| Australasian Journal of Educational Technology | 1,363 |
| International Journal of Computer-Supported Collaborative Learning | 1,717 |

In order to also consider the Journal Citation Reports Science Citation Index (JCR SCI), we repeated the iterative double check process with the journals indexed in the JCR SCI and obtained another list of journals, namely ET-FL-JCR-SCI list. Table 3 shows the first four journals of this list that corresponds to the journals from the JCR-SCI selected for this review. At this point we decided to include in the review, studies published in the first 4 journals of each list (ET-FL-JCR-SCI and ET-FL-JCR-SSCI). However, the “Internet and Higher Education” journal was not considered in the review since does not have studies published about AR in education. As a result, we included one additional journal from the ET-FL-JCR-SSCI list so that the number of journals considered can be equal. Those journals are the most relevant journals in Educational Technology according to our analysis. Those results were validated by comparing them with the SJR and SNIP indexes obtaining similar results.
Table 3. List of the first 4 journals of “ET-FL-JCR-SCI list”

<table>
<thead>
<tr>
<th>Journal title</th>
<th>Impact factor (JCR SCI 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge-based systems</td>
<td>4,104</td>
</tr>
<tr>
<td>Expert systems with applications</td>
<td>1,854</td>
</tr>
<tr>
<td>IEEE Transactions on education</td>
<td>0,95</td>
</tr>
<tr>
<td>IEEE Intelligent Systems</td>
<td>1,93</td>
</tr>
</tbody>
</table>

Step 1b: Inclusion and exclusion criteria

Taking into account the research questions, we considered general criteria that define the time frame for the study and the type of studies that are relevant. Accordingly, we defined the following criteria:

General Criteria:
- Studies published between 2003 and 2013.
- Studies that describe applications or frameworks for augmented reality in education.

Specific Criteria:
- Studies that report advantages, disadvantages, affordances, limitations, features, uses, challenges and effectiveness of augmented reality in educational settings.
- Studies that describe applications considering a user model and/or adaptive processes combined with augmented reality.
- Studies that describe applications of augmented reality in education for people in contexts of diversity.
- Studies describing the evaluation methods for augmented reality applications in educational scenarios.

The following exclusion criteria were defined and accordingly, studies meeting these criteria were excluded:
- Studies not identified as “Articles” in the journals selected (e.g., book reviews, books, editorial publication information, book chapters, etc.).
- Studies that mention the term “augmented reality” but are actually about virtual reality or other topics (and the term appears only in the references section).

Step 1c: Categories for the analysis and data coding

In this step, we defined a group of categories of analysis with their corresponding sub-categories according to each research question. Categories help us in grouping studies according to their shared characteristics. During the systematic review process, some sub-categories emerged and others were refined in order to cover all emerging information. The list of categories for the analysis classified by research questions (RQ) is as follows:

RQ1 - What are the uses, purposes, advantages, limitations, effectiveness and affordances of augmented reality in educational settings?
- Reported purposes of using AR.
- Reported advantages of AR.
- Reported limitations of AR.
- Reported effectiveness of AR.
- Type of AR.

RQ 2 - Have the inclusion of combined adaptive or personalized processes been considered in augmented reality applications?
- Type of adaptation process.
- Type of user modeling.

RQ 3 - How has augmented reality addressed the special needs of access and people preferences in educational settings?
- Special Need addressed.
- Intervention method.
RQ 4 - What are the evaluation methods considered for augmented reality applications in educational scenarios?

- Research sample.
- Research method
- Time dimension.
- Data collection method.

Content analysis allows to find the research trends of a topic by analyzing the articles’ content and grouping them according to the shared characteristics (Hsu, Hung, & Ching, 2013). This method was applied in order to extract the information of each paper. Two of the authors of the paper manually coded the studies separately according to their characteristics and classified them according to the categories and sub-categories defined. In case of discrepancy, the coders resolved it through discussion.

Results (Steps 2 and 3)

In this section the results of conducting the review are described and discussed. In step 2a we searched manually in the selected journals and applied the inclusion and exclusion criteria in order to select the studies for the review. As a result of this process we selected 32 studies from journals. Steps 2b and 2c were carried out by reading the papers completely and the data coding process was performed taking into account the categories defined in step 1c. In order to present the results this section was organized taking into account each research question addressed.

In total 30 studies were analyzed from the 5 journals selected from the JCR-SSCI and 2 studies were analyzed from the 4 journals selected from the JCR-SCI. Table 4 shows the number of studies analyzed by journal. It is important to note that in the table, the year 2013* includes the papers published until February 2014.

By analyzing the year of publication of the studies considered we found that the number of published studies about AR in education has progressively increased year by year specially during the last 4 years. This means that many researchers are interested in exploring the features, advantages, limitations of AR in educational settings. According to these results, AR in education is an emerging topic and this finding corroborates the ideas of Wu, Lee, Chang, & Liang (2013) and Cheng & Tsai (2012), who point out that the research on AR in education is in the initial phase. As Bujak et al. (2013) suggest: “Augmented reality (AR) is just starting to scratch the surface in educational applications.” One of the issues that emerge from these findings is that more research needs to be undertaken in the topic of AR in education.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Studies analyzed (2003-2013*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCR-SSCI Journals</td>
<td>Total: 30</td>
</tr>
<tr>
<td>Computers &amp; Education</td>
<td>23</td>
</tr>
<tr>
<td>Internet and Higher Education</td>
<td>0</td>
</tr>
<tr>
<td>British Journal of Educational Technology</td>
<td>4</td>
</tr>
<tr>
<td>Australasian Journal of Educational Technology</td>
<td>1</td>
</tr>
<tr>
<td>International Journal of Computer-Supported Collaborative Learning</td>
<td>2</td>
</tr>
<tr>
<td>JCR-SCI Journals</td>
<td>Total: 2</td>
</tr>
<tr>
<td>Knowledge-based systems</td>
<td>0</td>
</tr>
<tr>
<td>Expert systems with applications</td>
<td>1</td>
</tr>
<tr>
<td>IEEE Transactions on education</td>
<td>1</td>
</tr>
<tr>
<td>IEEE Intelligent Systems</td>
<td>0</td>
</tr>
</tbody>
</table>

In the following subsections, our findings with respect to each research question are presented.

What are the uses, purposes, advantages, limitations, effectiveness and affordances of augmented reality in educational settings?

With respect to the uses of AR in education, Table 5 presents the results obtained from the data coding process in the category of “Field of education.” This table clearly shows the use of augmented reality by each field of education.
The most striking result to emerge from the data is that most of the studies (40.6%) were applied in the field of “Science.” This result indicates that most of the research done in AR applied to education has been concentrated on identifying the benefits of AR in science education. A possible explanation of this is that AR has demonstrated to be effective when applied to lab experiments (Ibáñez, Di Serio, Villarán, & Delgado Kloos, 2014; Lin, Duh, Li, Wang, & Tsai, 2013; Enyedy, Danish, Delacruz, & Kumar, 2012), ecology (Wrzesien & Alcañiz Raya, 2010), field trips (Kamarainen et al., 2013), mathematics and geometry (Blake & Butcher-Green, 2009), scientific issues (Chang, Wu, & Hsu, 2013) and in general, activities where students can see things that could not be seen in the real world or without a specialized device. Besides that, students “do not have to use their imagination to envision what is happening. They can see it” (Furió, González-Gancedo, Juan, Seguí, & Rando, 2013) which also means that AR is effective for teaching abstract or complex concepts. A prior study has noted the importance of AR in science education. (Cheng & Tsai, 2012).

Following “Science” learning, “Humanities & Arts” was the second field of education in which AR was applied the most (21.9%). Studies in this field of education focused on language learning (Liu & Tsai, 2013; Chang, Lee, Wang, & Chen, 2010; Ho, Nelson, & Müeller-Wittig, 2011; Liu & Chu, 2010), visual art and painting appreciation (Di Serio, Ibáñez, & Kloos, 2013; Chang et al., 2014), and culture and multiculturalism (Furió et al., 2013). Interestingly, AR has been widely used in language learning due to the possibility of augment information and combining it with contextual information to provide new experiences in language learning. On the other hand, thanks to the possibility of adding virtual information to the real world AR has been applied in painting appreciation in order to provide an enhanced experience.

In “Social Sciences, Business and Law” and “Engineering, manufacturing and construction,” AR is being explored. Only 12.5% of the studies reviewed applied AR in “Social Sciences” and 15.6% applied AR in Engineering, manufacturing and construction.

Finally the results of our review show that the less explored fields of education are “Health and welfare” (3.1%) and Services and Others (travelling, transport, security services and hotel) with 6.3% of the studies reviewed. According to our review, no investigations have delved in the field “Educational” (teacher training in all levels of education) as well as the field of agriculture. The present results are significant in order to encourage researchers to explore the use of AR in teacher training and agriculture, forestry, fishery, veterinary, etc.

<table>
<thead>
<tr>
<th>Table 5. Augmented reality uses by “Field of education”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-category</td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Educational</td>
</tr>
<tr>
<td>Humanities &amp; Arts</td>
</tr>
<tr>
<td>Social Sciences Business and Law</td>
</tr>
<tr>
<td>Science</td>
</tr>
<tr>
<td>Engineering, manufacturing and construction</td>
</tr>
<tr>
<td>Agriculture</td>
</tr>
<tr>
<td>Health and welfare</td>
</tr>
<tr>
<td>Services and Others</td>
</tr>
</tbody>
</table>

Regarding the “Target group,” this category refers to the level of education of participants in the experiments in which the study of AR in education was carried out. Table 6 summarizes the results. This table is quite revealing in several ways. First, it is worth noticing that AR has been mostly applied in higher education settings (Bachelor’s or equivalent level) and compulsory education (primary, lower and upper secondary education). Most of the studies reviewed in these target groups applied AR for motivating the students, explaining topics, adding information and other purposes that are discussed later. It seems possible that AR has been applied in settings with this target group in order to improve the educational experience of the students and motivate and engage them by taking advantage of the features of this technology. In the studies reviewed there were no evidence of AR applications in the field “Early childhood education” (0%). A possible explanation of this result is that the technology could not be ready for being used by children since many aspects of interaction, such as the tracking and use of markers, need to be solved. We encourage researchers to explore the use of AR in this field.

On the other hand, “Post-secondary non-tertiary education” (0%) and “Short-cycle tertiary education” (3.1%) are target groups that need further research on the impact of AR in educational settings. This target groups are part of the
Vocational Educational Training (VET) in which AR could provide benefits in the learning process for facilitating the access to the labor market. So far, not many studies have been reported in this area. Finally, there were no evidence of using AR in “Masters or equivalent level” (0%) and “Doctoral” (0%) educational settings. This result may be explained by the fact that Master’s and PhD students typically are involved in creating new AR applications for the other levels.

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early childhood education</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Primary education</td>
<td>6</td>
<td>18.75</td>
</tr>
<tr>
<td>Lower secondary education</td>
<td>6</td>
<td>18.75</td>
</tr>
<tr>
<td>Upper secondary education</td>
<td>4</td>
<td>12.50</td>
</tr>
<tr>
<td>Post-secondary non-tertiary education</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Short-cycle tertiary education</td>
<td>1</td>
<td>3.13</td>
</tr>
<tr>
<td>Bachelor’s or equivalent level</td>
<td>11</td>
<td>34.38</td>
</tr>
<tr>
<td>Master’s or equivalent level</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Doctoral</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Informal Learning</td>
<td>2</td>
<td>6.25</td>
</tr>
<tr>
<td>Not mentioned in the study</td>
<td>2</td>
<td>6.25</td>
</tr>
</tbody>
</table>

Table 6. Target group in which AR studies were carried out

With respect to category “Purposes of using AR” in education, table 7 summarizes the results. Since one study can report more than one purpose, each study can meet more than one sub-category. It can be seen from this data that most of the studies used AR with the purpose of explaining a topic (43.7%) and augment information (40.6%). Explaining the topic refers to the use of an AR application in order to support the learning of a specific topic (Wrzesien & Alcañiz Raya, 2010; Chang et al., 2013). On the other hand, “augment information” refers to the use of AR for providing supplemental material by means of markers placed on printed material that students used to access digital resources (Huang, Wu, & Chen, 2012; Chen, Teng, & Lee, 2011).

Table 7 also shows that the purposes of using AR combined with “Educational Game” (18.7%) and for “Lab experiments” (12.5%) are being explored. In this sense, we encourage researchers to explore in detail the uses of AR in educational games in order to identify its features, advantages and drawbacks. Furió et al., (2013) claim that “there are few mobile learning games that use this technology.” Further research regarding to the role of AR for supporting lab experiments needs to be done, for example, the analysis of the impact of AR for reducing the cost of lab experiments or its strengths for offering a most inclusive experience for people with disabilities. Furthermore, according to the results, very little was found in the literature on using AR for activities for “Exploration” and discovering the world through AR (3.1%) and no studies were found with focus on using AR for evaluating a topic (0%) and the use of AR for other educational purposes (0%) different from the ones mentioned before.

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explaining the topic</td>
<td>14</td>
<td>43.75</td>
</tr>
<tr>
<td>Evaluation of a topic</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Lab experiments</td>
<td>4</td>
<td>12.50</td>
</tr>
<tr>
<td>Educational Game</td>
<td>6</td>
<td>18.75</td>
</tr>
<tr>
<td>Augment information</td>
<td>13</td>
<td>40.63</td>
</tr>
<tr>
<td>Exploration</td>
<td>1</td>
<td>3.13</td>
</tr>
<tr>
<td>Other purposes</td>
<td>0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 7. Purposes of using AR in educational settings

Another category analyzed in this systematic literature review deals with the “Reported Advantages” of AR in educational settings. Table 8 shows the results of the reported advantages identified in the studies analyzed. Since one study can report more than one advantage, each study can meet more than one sub-category. From the results, it can be seen that the major advantages reported in the studies are: “Learning gains” (43.7%) and “Motivation” (31.2%). These results corroborate the benefits of AR for improving the learning performance and motivating students (Liu & Chu, 2010; Di Serio et al., 2013; Jara et al., 2011; Chang et al., 2014). Some studies have reported other advantages of AR that are listed in table 8. However, these advantages need to be further explored in order to understand the real benefits of AR-based learning experiences. On the other hand, very little was found in the
literature on advantages of AR in educational settings such as: “Increase capacity of innovation” (6.2%), “creating positive attitudes” (6.2%), “Awareness” (3.1%), “Anticipation” (3.1%), “Authenticity” (3.1%), and “Novelty of the technology” (0%). In this sense, there is a need of more research in order to validate if those factors are advantages of AR in education.

Although interaction and collaboration have emerged as main advantages of AR, surprisingly data collection methods (discussed later in research question 4) such as focus groups or conversational analysis did not appear during the review. There are many evaluation mechanisms that have not been explored because the technology is not enough mature, so there is a gap between the affordances of AR, its advantages, uses, research methodologies and the evaluation mechanisms applied.

| Table 8. Reported advantages of AR in educational settings |
|---------------------------------------------|-----------------|-----------------|
| Sub-category                                | Number of studies | Percentage (%)  |
| Learning gains                              | 14              | 43.75           |
| Motivation                                  | 10              | 31.25           |
| Facilitate Interaction                      | 5               | 15.63           |
| Collaboration                               | 6               | 18.75           |
| Low cost                                   | 4               | 12.50           |
| Increase the experience                     | 4               | 12.50           |
| Just-In-time Information                    | 4               | 12.50           |
| Situated Learning                          | 3               | 9.38            |
| Student-centred                            | 3               | 9.38            |
| Students’ attention                         | 3               | 9.38            |
| Enjoyment                                  | 3               | 9.38            |
| Exploration                                | 4               | 12.50           |
| Increase capacity of innovation            | 2               | 6.25            |
| Create positive attitudes                  | 2               | 6.25            |
| Awareness                                  | 1               | 3.13            |
| Anticipation                               | 1               | 3.13            |
| Authenticity                               | 1               | 3.13            |
| Novelty of the technology                  | 0               | 0.00            |

Turning now to the category “Limitations of AR”, this category aims to identify the limitations of AR in educational settings. Results are shown in Table 9. From this data it can be seen that the most reported limitation in the studies reviewed are “Difficulties maintaining superimposed information” (9.3%). Students may feel frustrated if the application does not work properly or if it is difficult for them to use the markers or the device in order to see the augmented information. In order to overcome this limitation there is a need of improving the algorithms for tracking and image processing. In addition to this, it is recommended that further research be undertaken in usability studies for AR applications in education as well as guidelines for designing AR-based educational settings.

Another limitation reported was “Paying too much attention to virtual information” (6.2%). This limitation is related to the novelty of this technology when it is used for the first time in the classroom. So, students may be distracted by the virtual information showed or the technology itself. “Intrusive Technology” (6.2%) was also a limitation reported which is connected with the use of HDM (Head-mounted displays) (Zarraonandia, Aedo, Díaz, & Montero, 2013) because the device can interrupt the natural interaction with others.

Other limitations reported in the studies are: “Designed for a specific knowledge field” (3.3%) and “Teachers cannot create new learning content” (3.1%). In this sense, it is recommended that further research be undertaken in authoring tools for creating AR activities so that teachers can create their own content with AR support.

| Table 9. Limitations of AR in educational settings |
|---------------------------------------------|-----------------|-----------------|
| Sub-category                                | Number of studies | Percentage (%)  |
| Designed for a specific knowledge field     | 1               | 3.13            |
| Teachers cannot create new learning content | 1               | 3.13            |
| Difficulties maintaining superimposed information | 3           | 9.38            |
| Paying too much attention to virtual information | 2          | 6.25            |
With respect to the category “Effectiveness of AR,” table 10 shows the results. Since one study can report more than one sub-category of effectiveness, each study can meet more than one sub-category. Most of the studies reported that AR applications lead to “Better learning performance” (53.3%) in educational settings. “Learning motivation” (28.1%) and “Student engagement” (15.6%) were also reported. The results show that AR is a promising technology for improving the student’s learning performance and motivate the students to learn thanks to the interaction and graphical content used. “Improved perceived enjoyment” (12.5%) and “Positive attitudes” (12.5%) were less reported but are also important in educational settings.

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better learning performance</td>
<td>17</td>
<td>53.13</td>
</tr>
<tr>
<td>Learning motivation</td>
<td>9</td>
<td>28.13</td>
</tr>
<tr>
<td>Improve perceived enjoyment</td>
<td>4</td>
<td>12.50</td>
</tr>
<tr>
<td>Decrease the education cost</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Positive attitudes</td>
<td>4</td>
<td>12.50</td>
</tr>
<tr>
<td>Student engagement</td>
<td>5</td>
<td>15.63</td>
</tr>
</tbody>
</table>

Regarding the “Type of AR” considered in the studies reviewed, table 11 summarizes the results. We considered three types of AR according to the classification of Wojciechowski & Cellary (2013): marker-based AR, marker-less AR and location-based AR. Marker-based AR is based on the use of markers. Markers are labels that contain a colored or black and white pattern that is recognized or registered by the AR application through the camera of the device in order to fire an event that can be, for instance, to show a 3D image in the screen of the device located in the same position where the marker is. Marker-less AR is based on the recognition of the object’s shapes. And location-based AR superimposes information according to the geographical location of the user.

The results in table 11 reveal that most of the studies used “Marker-based AR” (59.3%) which means that most of the applications developed for educational settings use markers. A possible explanation for this result is that currently the tracking process of markers is better and more stable compared to the marker-less tracking techniques. The use of static markers decrease the tracking work needed and reduce the number of objects to be detected (El Sayed et al., 2011). Therefore for educational settings the use of markers could be recommended so that students can have a better experience with the technology until better techniques for tracking can be developed for marker-less AR. “Marker-less AR” has not been widely used in educational settings (12.5%). However, there is a trend of using Microsoft Kinect sensors and similar technologies in order to create AR applications for educational settings (Fallavollita et al., 2013; Pillat, Nagendran, & Lindgren, 2012). Microsoft Kinect provides some advantages in tracking and registering objects in marker-less AR.

<table>
<thead>
<tr>
<th>Code</th>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC1</td>
<td>Marker-based AR</td>
<td>19</td>
<td>59.38%</td>
</tr>
<tr>
<td>SC2</td>
<td>Marker-less AR</td>
<td>4</td>
<td>12.50%</td>
</tr>
<tr>
<td>SC3</td>
<td>Location-based AR</td>
<td>7</td>
<td>21.88%</td>
</tr>
<tr>
<td>SC4</td>
<td>Not specified in the study</td>
<td>2</td>
<td>6.25%</td>
</tr>
</tbody>
</table>

Interestingly, the development of “Location-based AR” (21.8%) applications is major compared to marker-less AR applications. This can be due to the availability of sensors in mobile devices like the accelerometer, gyroscope, digital compass and the possibility of using GPS. These technological advancements open possibilities for developing applications of AR that can be aware of the user’s location in order to show information according to the geographical position and/or orientation.
Have the inclusion of combined adaptive or personalized processes been considered in augmented reality applications?

In the studies reviewed only 2 out of 32 studies report some kind of personalized process and 1 out of 32 considered a user modeling process. Barak & Ziv (2013) created “Wandering” which is an application for creating location-based interactive learning objects (LILOs) and considers personalization as an “important requirement of the 21st century skills” (Barak & Ziv, 2013). Personalization is considered for meeting the needs and interests of the individual learners. However, in the study where the authors describe the Wandering application is not clear if they have a user model. On the other hand, Blake & Butcher-Green (2009) propose an application for customized training based on a scaffolding instructional approach and an agent architecture in order to training individuals from diverse backgrounds. The type of adaptation process considered by the application is personalization based on historical training profiles. However in the paper is not clear if the information for the user model comes from the learner’s profile. In addition to this, the authors states that the system was being integrated with the AR environment when the paper was written. The results of the paper are based on a simulated AR environment (Blake & Butcher-Green, 2009).

How has augmented reality addressed the special needs of access and people preferences in educational settings?

In the studies reviewed from journals there was no evidence of AR applications in educational settings that address the special needs of students. This finding corroborates the idea of Wu, Wen-Yu, Chang, & Liang (2013) who state that few systems have been designed for students with special needs. According to Lindsay (2007) the opportunities for children with special needs and disabilities can be improved by a major policy initiative called “inclusion”. Inclusive education is more than integration because integration refers to the learner adapting to the educational setting while inclusion means that the educational setting adapts to the learner in order to meet their needs (Lindsay, 2007). Within this sense AR may offer unique advantages and benefits in order to create inclusive AR-based educational settings. Further research is needed in order to identify the effectiveness and advantages of AR applications for addressing the special needs of students.

What are the evaluation methods considered for augmented reality applications in educational scenarios?

With respect to the evaluation methods for AR applications in educational settings we considered four sub-categories for the analysis. The results show that, regarding to “Research Samples” (table 12), most of the studies used medium research samples “between 30 and 200” (78.1%) and some studies considered small research samples “30 or less than 30” (18.7%). In our review we did not find studies that used research samples greater than 200 participants (“More than 200” (0%)). A possible explanation of this result is that greater research samples would need more devices (handheld devices, PC, web cam, tablets, etc.) so that each participant can have one device.

Table 12. Research samples in the studies reviewed

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 or less than 30</td>
<td>6</td>
<td>18.75</td>
</tr>
<tr>
<td>Between 30 and 200</td>
<td>25</td>
<td>78.13</td>
</tr>
<tr>
<td>More than 200</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Not Specified in the study</td>
<td>1</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Regarding the “Research Methods,” table 13 shows that most of the studies applied “mixed methods” (46.8%), “Qualitative-Exploratory-Case study” (21.8%), “Quantitative-Descriptive research” (15.6%) and “Qualitative-Exploratory-Pilot Study” (12.5%) as research methods to conduct the study. Few studies have applied “Quantitative-Explanatory and Causal research” (3.1%) and “Qualitative-Exploratory-Experience Survey” (0%).

Table 13. Research methods applied

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative-Exploratory-Case Study</td>
<td>7</td>
<td>21.88</td>
</tr>
<tr>
<td>Qualitative-Exploratory-Pilot Study</td>
<td>4</td>
<td>12.50</td>
</tr>
<tr>
<td>Qualitative-Exploratory-Experience Survey</td>
<td>0</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Turning now to the “Time dimension” of the studies reviewed, table 14 shows that almost all of the studies were identified as “Cross-sectional” (93.7%) and only 6.2% of the studies were identified as “Longitudinal Study.” An implication of this result could be that the novelty of the technology in cross-sectional studies may affect the results since students can be engaged with the AR application because it is new for them. Future studies conducted as longitudinal studies need to be undertaken in order to follow the students in the long term and identify the advantages, benefits, limitations when students are exposed to this technology for a long period and also when students are used to using AR in the classroom as well as analyze the student’s behavior in different learning scenarios.

Table 14. Time dimension of the studies reviewed

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional Study</td>
<td>30</td>
<td>93.75</td>
</tr>
<tr>
<td>Longitudinal Study</td>
<td>2</td>
<td>6.25</td>
</tr>
</tbody>
</table>

Finally, for “Data Collection methods” as table 15 shows, most of the studies applied “Questionnaires” (75%), “interviews” (28.3%), “surveys” (18.7%) and “cases observation” (9.3%) as data collection methods. “Focus group” (0%) and “Writing Essay” (3.1%) have either not been used or used very little. Since one study can apply more than one data collection method this study counts for more than one category.

Table 15. Data collection methods

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Number of studies</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaires</td>
<td>24</td>
<td>75.00</td>
</tr>
<tr>
<td>Interviews</td>
<td>9</td>
<td>28.13</td>
</tr>
<tr>
<td>Focus-groups</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Survey</td>
<td>6</td>
<td>18.75</td>
</tr>
<tr>
<td>Cases observation</td>
<td>3</td>
<td>9.38</td>
</tr>
<tr>
<td>Writing Essay</td>
<td>1</td>
<td>3.13</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>3.13</td>
</tr>
</tbody>
</table>

**Trends and future vision**

This study presents a detailed systematic review of the state of the art in Augment reality as a promising technology for supporting technology-enhanced learning. In this section we discuss our main findings highlighting what we consider as the strongest future directions of research in this field.

**AR technology**

Marker-based AR as mentioned in the results section is the most used approach for supporting the development of AR learning experience, followed by the location-based AR. A possible explanation for this result is that currently the tracking process of markers is better and more stable compared to the marker-less tracking techniques. Besides that one of the advantages of marker-based AR is the facility of implementation due to the available libraries which support the development process. There is a challenge around the improvement of recognition algorithms for human forms as a promising feature in the process of achieving more immersive and not intrusive AR learning experiences (Zarraonandia et al., 2013). Accessibility and usability of the AR learning experiences are two important issues to be addressed in future research since few studies have reported research on this field. In fact, only 4 out of 32 studies consider those factors (Di Serio et al., 2013; Ibáñez et al., 2014; Ho et al., 2011; Cuendet et al., 2013). Further research need to be undertaken in usability studies for AR applications in education as well as guidelines for designing AR-based educational settings.
Some recent studies have reported new research directions:
- There is a need for “new methods for creating interactive 3D content for AR learning environments” (Wojciechowski & Cellary, 2013; K.-E. Chang et al., 2014) and creating authoring tools for teachers to create content.
- Understand how to design AR learning experiences according to the topic taking into account the skills of learners (Bujak et al., 2013).
- Creating multisensory experiences with AR (Ho et al., 2011) and explore their impact in the learning outcomes.
- Carry out more studies for understanding the user experience and knowledge construction processes in AR applications (Lin et al., 2013).

Attention to diversity and special needs in the learning process

Although there are not many reported papers considering samples which include people with special need of access to educational context, the multimodal possibilities of AR applications seems to be a good option for addressing the special necessities of diverse population.

Some challenges are the definition of frameworks for user model representations as well as frameworks to support personalization processes. In fact, a starting point could be the analysis of existing frameworks for AR (Bujak et al., 2013; Chen et al., 2011; Price & Rogers, 2004) in order to analyse the feasibility of integration of these frameworks to offer an augmented but also an adaptive learning experience to users. These frameworks should at least include the definition of semantics to represent the user’s profiles and their context, semantics to represent the metadata of the AR resources as well as semantic to represent different possibilities of adaptations.

On the other hand, the multimodal possibilities of AR applications seem to be a good option for supporting therapy processes for people with sensorial and physics impairments. It demonstrates that AR increases the motivation of the user developing specific tasks which is one of the challenges in therapy processes (Correa, Ficheman, Nascimento, & Lopes, 2009). Finally, it can thus be suggested that frameworks for personalized AR should consider pedagogical and didactical components in order to guide the development of AR-based educational settings.

The need for longitudinal studies

Based on the conducted analysis, we conclude that more studies need to be undertaken considering a large scale evaluation and longitudinal evaluations in future researches. Many cross-sectional studies have been conducted as shown in the results of this review, being an efficient research method in order to establish comparisons between AR learning experiences with respect to other cases of learning experience. However, in the case of educational settings is also important to study the evolution of knowledge and skills over time, as proposed by longitudinal studies. Long term analysis of the AR learning experience could give important inputs about the suitability of this technology for supporting significant learning (Mendoza Gómez, 2005).

Vocational educational training (VET) as target groups for future research

According to our research, only 3.1% of the studies were carried out considering a sample of students from vocational educational training institutions. From our point of view, VET institutions are promising research partners not only for validation but also for demonstrating the possibilities of AR learning scenarios for improving and acquiring professional competences. Possibilities offered for AR could reduce the cost of carrying out some learning experience where expensive learning material is necessary. For example, physical materials to learn how to create a gem could be difficult to buy by institutions with limited economic resources. In this sense, combining virtual objects as different kind of physical materials with real objects such as the students’ hands could be a good option to explore.
Conclusions

In this paper a systematic literature review was reported. In total 32 studies from journals were analyzed by using the content analysis method. We analyzed the following factors of the studies selected: Field of education, target group, type of AR, reported purposes, advantages, limitations, affordances and effectiveness of AR in educational settings. Besides that, adaptation processes and user modeling in AR as well as the addressing of individual special needs with AR applications was also considered for the analysis. Regarding the evaluation methods we analyzed the research sample, research method, and time dimension of the study. Furthermore, we defined a validated method for selecting journals through a methodologically strong and consistent process that can be applied for systematic reviews in other topics.

A short summary of the main findings of this review are:

- The number of published studies about AR in education has progressively increased year by year specially during the last 4 years.

- Science and Humanities & Arts are the fields of education where AR has been applied the most. Health & welfare, Educational (teacher training) and Agriculture are the research fields that were the least explored fields.

- AR has been mostly applied in higher education settings and compulsory levels of education for motivating students. Target groups like early childhood education and Vocational educational Training (VET) are potential groups for exploring the uses of AR in future.

- Marker-based AR is the most used type of AR. In addition location-based AR is being widely applied. This can be due to the availability of sensors in mobile devices like the accelerometer, gyroscope, digital compass and the possibility of using GPS. Marker-less AR needs some improvement in algorithms for tracking objects but the use of Microsoft Kinect is becoming more and more popular.

- The main purpose of using AR has been for explaining a topic of interest as well as providing additional information. AR educational games and AR for lab experiments are also growing fields.

- The main advantages for AR are: learning gains, motivation, interaction and collaboration.

- Limitations of AR are mainly: difficulties maintaining superimposed information, paying too much attention to virtual information and the consideration of AR as an intrusive technology.

- AR has been effective for: a better learning performance, learning motivation, student engagement and positive attitudes.

- Very few systems have considered the special needs of students in AR. Here there is a potential field for further research.

- Most of the studies have considered medium research samples (between 30 and 200 participants), and most of the studies have used mixed evaluation methods. The most popular data collection methods were questionnaires, interviews and surveys and most of the studies were cross-sectional.

This work contributes to existing knowledge in AR in educational settings by providing the current state of research in this topic. This research also has identified relevant aspects that need further research in order to identify the benefits of this technology to improve the learning processes.

Acknowledgements

This work is supported in part by the Spanish Science and Education Ministry in the ARrELS project (TIN2011-23930). Jorge Bacca, Silvia Baldiris and Ramon Fabregat belong to the BCDS group (ref. GRCT40) which is part of the DURSI consolidated research group COMUNICACIONES I SISTEMES INTELLIGENTS (CSI) (ref. SGR-1202).
Jorge Bacca would like to thank the financial support for the Grant (FPI-MICCIN) provided by the Spanish Ministry of Economy and Competitiveness (BES-2012-059846). The authors acknowledge the support of NSERC, iCORE, Xerox, and the research related gift funding by Mr. A. Markin.

References


Closing the Missing Links and Opening the Relationships among the Factors: A Literature Review on the Use of Clicker Technology Using the 3P Model

Jae Hoon Han
Project Management Office (PMO), McGill University, #1450-688 Sherbrooke Street west, Montreal, Canada, H3A 3R1 // jaehoon.han@mail.mcgill.ca

ABSTRACT
Clicker technology is one of the most widely adopted communication systems in college classroom environments. Previous literature reviews on clicker technology have identified and thoroughly documented the advantages, disadvantages, and implications of the use of this technology; the current review is intended to synthesize those earlier findings and recast them in terms of the interrelationship between the “3 Ps” of the 3P model: Presage, Process, and Product factors. Using this guided framework enables the identification of the most up-to-date trends and issues in clicker studies published in peer-reviewed journals since 2009. The review shows that recent clicker studies have examined the effects of clickers in terms of student presage factors (cognitive, non-cognitive, background factors), instructor presage factors (instructor effects and the level of the course taught), process factors (delivery method, instructional activities, and assessment and feedback), and product factors (cognitive and non-cognitive outcomes). A heat-mapping approach is used to facilitate the interpretation of the results. The findings also discuss missing/unaddressed links and the untapped relationships among instructional factors in these studies. This study concludes that teaching and learning with the use of clicker technology is a complex and relational phenomenon; factors that are currently under-explored should be examined using more rigorous research methods to close gaps in the literature and to enhance understanding of the use of clickers in classroom learning environments.

Keywords
Clickers, Pedagogical approaches with technology, Information visualization methods, Interactive learning environments

Introduction
Research into teaching and learning in higher education consistently seeks answers to the following two questions: What is the best delivery method for enhancing student learning outcomes? How does a specific instructional strategy make student learning effective and efficient in a large classroom setting? (e.g., Biggs, 2003; McKeachie, 1990). These questions have led scholars to seek the best-fitting factors and identify interrelationships among the factors that enhance the quality of student learning. One possible workaround to resolve these issues is to create a “student-centered learning environment,” often described as a classroom setting in which students actively engage and interact within the classroom and with various instructional activities. Ultimately, a range of factors can be identified that have the potential to concurrently enhance the quality of various student learning outcomes.

The Classroom Communication System (CCS; hereafter referred to as “clickers”) is one of the most widely adopted interactive technologies used in classroom instruction worldwide (White, Syncox, & Alters, 2011). Clicker systems comprise two main components: the response device and the receiver system. Instructors introduce questions in various formats (multiple choice or right/wrong answers) to the students via the classroom projector, and they respond to the questions using the response device (e.g., smartphone or number keypad). All the student responses are collected in the instructor’s receiver system embedded in the computer; instructors can also immediately post (project) the student response results back to the class in a few different visual formats. Thus, the use of clickers can provide immediate feedback even to a large number of students and reduce the time spent on the feedback cycle between instructors and students (Li, Chung, Chen, & Liu, 2009).

Studies (e.g., Caldwell, 2007) suggest that the potential benefits of clicker technology, including immediate feedback and assessment, may help to build student-centered classroom environments. More specifically, the use of clickers with appropriate pedagogical approaches may have the following positive impacts on teaching and learning in higher education: improved student involvement/engagement, clearer perceptions of learning/knowledge gains, and stronger affective learning outcomes (e.g., student attitudes). However, 2001 Physics Nobel prize recipient Carl Wieman suggests that the use of clicker technology can have deep and positive impacts on students only when the following specifications are met throughout: “...change in the classroom dynamic, the questions posed, and how they are followed up...and guided by an understanding of how people learn” (Wieman & Perkins, 2005, p. 39).
The specifications (or prerequisites) for improving student learning with clicker technology have been well documented in several literature reviews. For instance, Kay and LeSage (2009) reviewed 67 peer-reviewed journal articles and identify the following benefits and challenges of the use of clickers: assessment, classroom, learning benefits and technology-, teacher-, and student-based challenges. Since the most recently published literature review (Kay & LeSage, 2009) was done in 2009, it is important to review the clicker studies published after 2009 for the following reasons. First, in these studies, a wide range of new factors have been considered, including different groups of participants (e.g., multiple course levels; Dunn, Richardson, McDonald, & Oprescu, 2012), various types of interventions (e.g., different levels of clicker use; Elicker & McConnell, 2011), and instructional settings (e.g., different institutions; Lundeberg et al., 2011). These factors, however, have not been taken into account in any clicker literature reviews to date. Second, the most recently published literature review on this topic (Kay & LeSage, 2009) was conducted in 2009, and is, therefore, somewhat out of date. Third, previous literature reviews (e.g., Caldwell, 2007) were primarily interested in documenting the practical advantages of and barriers to adopting clickers into classrooms, rather than pursuing a concrete understanding of how the use of clickers is related to classroom instruction. Therefore, conducting new literature review is necessary and important to provide a more comprehensive understanding of this topic that has been published after 2009. All of these help educational technology practitioners and researchers to refresh and regain the most up-to-date information on the use of clickers and to synthesize new knowledge on clicker effects in relation to various instructional factors in classroom instruction.

This study aimed at (1) identifying the missing links within clicker studies in higher education, in order to arrive at a comprehensive understanding of research and development on clicker implementations and permit researchers and practitioners to work toward closing existing gaps, and (2) providing up-to-date theoretical and practical implications and issues for implementing clicker technology across academic disciplines in higher education by uncovering the various types of interrelationships among the factors. Because this work was conducted using the Presage-Product-Process (3P) Model (Biggs, 2003), the approach taken in the current review is deductive rather than inductive: the 3P model categorizes each identified construct as one of the 3P components and then synthesizes findings by focusing on the relationships among the factors examined in the studies.

The current review is organized as follows. First, a description of the criteria and selection process for retrieving articles from academic databases is provided and the guiding framework, the 3P model, is presented. The terms presage, process, and product are defined, and their possible interrelationships discussed. Next, results synthesized from the study review process are presented in terms of the 3P factors, using a heat-mapping information visualization method. Finally, practical implications for the design and implementation of clicker systems, together with implications for ongoing research, are presented in the conclusion.

Method

Literature search and selection approaches

The design of the current literature review was guided by recommendations and comments found in Boote and Belie (2005). Forward- and backward-citation tracking methods were employed to find articles on the use of clickers within the following academic databases: EBSCO, ERIC, ProQuest, PsycINFO, Scopus, and Google Scholar. In addition, Bruff’s (2014) bibliography of clicker studies was used to decrease the probability of articles being overlooked.

The following criteria were used to select the articles for review in the current study: (1) the articles were published in peer-reviewed journals between 2010 and 2013; (2) the studies examined more than one class session in a course as a unit of analysis (Trigwell, 2010) in college classroom instruction; (3) the studies examined the use of clickers in terms of instruction rather than other constructs (e.g., low-income mothers; Ginter, Maring, Paleg, & Valluri, 2013); (4) the studies provided descriptions of research design, data collection procedures, and results, rather than non-evidence-based pedagogical recommendations. Thus conference proceedings and institutional reports were not considered in this literature review.
In order to identify appropriate articles based on the criteria presented above, the following keywords and their combinations (AND and OR functions) were used during the retrieval process: Student*, Audience*, Class*, Comm*, Vot*, Response*, System*, and Clicker*.

Framework of the study

Objectives of the review are to (a) provide the most up-to-date information about the trends and issues in current research on the use of clickers in college instruction, and (b) identify implications for design and implementation of clickers for faculty, faculty developers, and educational technology staff who are involved in the use of clicker technology in higher education. The Presage-Process-Product (3P) model (Biggs, 2003) was adopted to guide the review process.

Framework

Using the 3P model as a framework facilitates the literature review process along several dimensions. First, the 3P model has been empirically examined and validated for virtually all academic disciplines worldwide (Biggs, 2003); as a result, this model can provide a more valid and reliable understanding of the relationship between the use of clickers and teaching and learning in higher education. Second, the 3P model emerged from context-specific studies on teaching and learning and therefore it conforms to the current study’s objective of understanding the use of clickers in college instruction. Third, the 3P model can provide a comprehensive understanding of the characteristics of the relationships among the constructs (i.e., presage, process, and product factors) of instruction with clickers. In other words, the 3P model framework can easily identify the relationship among or between the constructs, characterize the nature of those relationships, and provide a clear picture of how the constructs interact in the context of clickers use.

![Figure 1. Adopted framework from Biggs’ (2003) Presage-Process-Product (3P) model](image-url)


3P model

The framework (see Figure 1) was developed using the 3P model. The underlying assumptions of the 3P model are based on an understanding of teaching and learning as relational phenomena (Trigwell, 2010) that concurrently intertwine or interact with presage, process, and product factors in the learning environment as a closed system.

“Presage factors” are individual characteristics (of the student; of the instructor) that simultaneously affect and are affected by the process and product. Student presage factors include each individual student’s prior cognitive (e.g., knowledge) and non-cognitive (e.g., motivation) factors, which may or may not impact their learning processes and outcomes. Instructor/learning environment presage factors simultaneously affect student presage factors, learning processes and outcomes. The following themes are identified as instructor presage factors: course level, instructor/institutional effects, and pedagogical training.

“Process factors,” for the purpose of this study, include both teaching and learning processes and the interaction between student approaches to learning and instructor approaches to teaching. Student approaches to learning can be viewed on a continuum from deep to surface approaches. Whether a student’s approach to learning is deep or surface depends on the method of learning and the learning-focused activities that students adopt and practice. Although there is considerable variation among instructors’ approaches to teaching, they tend to fall somewhere on a continuum from teacher-centered to student-centered. In teacher-centered approaches, instructors transmit information in a didactic way, while using student-centered approaches, instructors adopt instructional strategies that achieve qualitative changes in student conceptions (i.e., active learning strategies). In this review, process factors are used to identify what instructional activities and strategies are adopted when clicker technology is implemented in college classroom instruction.

"Product factors” are learning outcomes that arise through the interaction between presage factors and processes in the 3P model. Specifically, two types of learning-outcome “products” are included in this study: cognitive outcomes (e.g., GPA) and non-cognitive outcomes (e.g., engagement). In most studies, cognitive and non-cognitive learning outcomes are among those most frequently assessed in the evaluation of clicker technology (Han & Finkelstein, 2013). In the current review, product factors are used to identify the measures (learning outcomes) that each clicker-technology study used to assess or evaluate student learning.

Using the 3P model to review clicker-use studies will help us hone in on several components of these studies: (a) what aspects of instructor and student characteristics were considered when implementing clicker technology; (b) how instructional strategies and activities were designed and implemented to achieve instructional goals; (c) what learning outcomes the authors of each study sought to assess or evaluate.

Analysis of studies and guiding protocols

The analysis of the clicker studies and the guiding protocols of the current review are based on Biggs (2003) and Kane, Sandretto, and Heath (2002). The 3P model (Biggs, 2003) was used to identify the research focus of each clicker-use study and the relationships among the various research foci. Kane et al.’s (2002) guiding protocols were used to investigate the relationships among the research methods, data collection, and implications for the use of clickers in college classroom instruction. The following information is presented for each reviewed study on the supplements website (http://mapping-2014.weebly.com/): author(s) and year, research focus, number of participants, presage, process and product factors, research methods, references, etc.

Visualization approaches

Heat-mapping approaches were used to visualize the results of the literature review using tableau public (Version 8.2). The heat-mapping approach that uses an area-based visualization technique with multi-way contingency tables to show the frequency distributions of hierarchical data in relation to the complete data set. According to Auber (2013), these mapping approaches can provide snapshots that assist readers to intuitively understand the data and results. The heat-mapping technique (Wilkinson & Friendly, 2009) was used to provide an appropriate visual framework for contrasting and comparing the co-occurrences among the factors.
Results and discussion

Results of literature search and selection approaches

Of over 200 articles retrieved from the academic database search, 84 articles were selected to be reviewed using the 3P model. 78 of these 84 articles were carried out in a discipline-specific educational setting (e.g., Physics), and only four studies were conducted in a multiple-discipline setting. Over 80% of the studies were carried out in North America; the majority (75.90%) of the studies used quantitative methodology; over 55% of the studies were published in 2012 and 2013. Interactive-mapping (see Figure 2) is available on the website (http://mapping-2014.weebly.com/).
Presage student factors

Several studies have examined student presage factors in response to the use of clickers in classroom instruction. These factors are mostly related to Cognitive, Non-Cognitive and Individual Background aspects of students (see Table 1).

Table 1. Description and examples of presage student factors

<table>
<thead>
<tr>
<th>Presage Student Factors (Frequency)</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive factors ($f = 8$)</td>
<td>Factors that evaluate student characteristics related to prior performance, metacognition (e.g., grade and academic ability).</td>
<td>Brady, Seli, &amp; Rosenthal (2013a,b), Freeman, Haak, &amp; Wenderoth (2011), Roth (2012), Sprague &amp; Dahl (2010), Trew &amp; Nelsen (2012)</td>
</tr>
<tr>
<td>Background factors ($f = 12$)</td>
<td>Factors pertaining to student demographic characteristics (e.g., campus location and gender)</td>
<td>Bright, Reilly Kroustos, &amp; Kinder (2013), Cline, Parker, Zullo, &amp; Stewart (2012), Dunn et al. (2013), Smith, Wood, Krauter, &amp; Knight (2011)</td>
</tr>
</tbody>
</table>

Cognitive factors

Eight studies examined the effects of clicker use on several cognitive factors associated with student learning: student metacognition, cognitive load, student prior achievement, and academic ability. Several of these studies examined the relationship between aspects of metacognition and student learning. Brady et al. (2013a, b) found that increased clicker use was closely related to increased levels of student metacognition, which successively affected student learning outcomes (measured by quiz scores). Jones et al. (2012) found that clicker-driven changes in students’ metacognition (i.e., regulation of cognition) were closely related to gender differences. Other studies have examined the relationships between academic ability and performance level, on the one hand, and student learning in response to the use of clickers, on the other (see Table 1). The results of these studies indicate that the use of clickers might be more effective with lower-performing students.

Non-cognitive factors

Thirteen studies have examined the relationships between the following non-cognitive factors and student learning with clicker technology (see Table 1): students’ academic emotions (e.g., Filer, 2010), students’ attitudes toward technology use (e.g., Dallaire, 2011), and students’ apprehension regarding communication (e.g., Fortner-Wood, Armistead, Marchand, & Morris, 2013). In Sutherlin et al.’s (2013) study, student attitudes toward clicker use did not significantly change over the course of the semester; however, other studies have consistently found a positive relationship between students’ attitudes, perceived engagement, learning (e.g., Evans, 2012), and final grades (e.g., Dallaire, 2011).

Background factors

Even though most studies examined some aspect of background factors for pre-test screening, only 12 studies focused on the relationship between background factors and student learning. Studies examining gender effects on learning with clickers have found mixed results. Two studies (Jones et al., 2012; Kang et al., 2012) found that female students were more likely than male students to favor clicker-based cooperative instruction, while other studies (e.g., Dunn et al., 2013) did not find gender effects. However, Trew and Nelsen (2012) examined the interaction between two primary background factors (gender and year of enrollment) and course grades, it was reported that these
Presage factors taken together predicted positive student-perceived learning and outcomes in clicker-based instruction.

Presage instructor factors

Table 2. Description and examples of presage instructor factors

<table>
<thead>
<tr>
<th>Presage Instructor Factors (Frequency)</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course level ($f = 3$)</td>
<td>Factors related to the context in which the clickers were implemented.</td>
<td>FitzPatrick, Finn, &amp; Campisi (2011), Lee, Garcia, &amp; Porter (2013)</td>
</tr>
<tr>
<td>Instructor/Institutional effects ($f = 5$)</td>
<td>Factors that explain the broader contextual factors (e.g., the relationship between the different institutions and the clicker effects on instruction).</td>
<td>Cline, Zullo, Duncan, Stewart, &amp; Snipes (2013), Emenike &amp; Holme (2012), Lundeberg et al. (2011)</td>
</tr>
<tr>
<td>Pedagogical training ($f = 1$)</td>
<td>Factors pertaining to instructors’ preparedness and knowledge about teaching with the use of clickers.</td>
<td>Han &amp; Finkelstein (2013)</td>
</tr>
</tbody>
</table>

Although presage instructor factors have not often been examined in relation to the use of clickers, several studies have reported interesting relationships among instructor factors, instructor perceptions, and student learning (see Table 2). Emenike and Holme (2012) found a significant difference in instructors’ perceived efficacy with clicker-based pedagogy across types of institutions and suggested that this difference may arise because instructors at doctoral universities are more likely to teach large-enrollment courses. Thus, it may be class size rather than institution type that is truly affecting the efficacy in this context. Han and Finkelstein (2013) also found that individual instructors’ pedagogical training experience was positively related to student-perceived engagement and learning. However, no significant relationship between course-level and student learning was found in the clicker-use studies (FitzPatrick et al., 2011).

Process factors

Table 3. Description and examples of process factors

<table>
<thead>
<tr>
<th>Process Factors (Frequency)</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery methods ($f = 25$)</td>
<td>Factors that compare the effectiveness of clicker use versus other delivery methods (e.g., raising hands and lecturing)</td>
<td>Bachman &amp; Bachman (2011), Evans (2012)</td>
</tr>
<tr>
<td>Instructional activities/Course design ($f = 45$)</td>
<td>Factors that examined the effects of clicker use on students regarding the implementation of various types of instructional strategies and course design (e.g., specific types of clicker questions)</td>
<td>Anthis (2011), Bunce, Fiens &amp; Neiles (2010), Dallaire (2011), Elicker &amp; McConnell (2011), Gray, Owens, Liang, &amp; Steer (2012)</td>
</tr>
<tr>
<td>Assessment and feedback ($f = 8$)</td>
<td>Factors that examined a high- or low-stakes testing and methods of evaluating student responses in relation to the use of clickers (e.g., formative/summative assessment and feedback)</td>
<td>Chen &amp; Lan (2013), Gachago, Morris, &amp; Simon (2011), James &amp; Willoughby (2011)</td>
</tr>
</tbody>
</table>

Delivery method

Although use of clicker technology (in comparison to non-use) was generally found to have a positive outcome on student non-cognitive learning outcomes (Caldwell, 2007), some studies did not report any significant effects of clickers over any other type of delivery method (e.g., clickers, flashcards, and raising hands); Elicker & McConnell, (2011) reported a negative relationship between clicker use and student exam scores (e.g., Anthis, 2011). The current study supports the finding of other literature review articles (e.g., Kay & LeSage, 2009) that the use of clicker
technology is related to students’ positive perceptions of their engagement and learning; however, it remains unclear whether this positive perception correlates with enhanced student performance.

### Instructional activities with clickers

Several studies explored the effects of different instructional strategies and course design approaches on student learning in response to the use of clickers (e.g., Levesque, 2011). These studies primarily examined teaching approaches with clickers (e.g., Jones et al., 2012), design of clicker questions (e.g., Hogan & Cernusca, 2013), and ways of representing questions using clickers (Perez et al., 2010). The results of these studies remain mixed; Rush et al. (2013) did not find any significant difference in student learning performance regardless of the use of peer-instruction and individual-knowledge clicker questions, while Smith et al. (2011) found that peer-instruction approaches (with instructor explanation) were more closely related to overall student performance than any other methods using clickers.

### Assessment and feedback approaches with clickers

The merits of formative versus summative assessment and feedback approaches are still under debate. White et al. (2011) suggested that a policy of awarding grade points for student clicker answers would be inappropriate due to the possibility of cheating; Han and Finkelstein (2013) confirmed White et al.’s (2011) assessment that instructors’ use of formative feedback was positively related to student-perceived engagement and learning (whereas summative feedback was not). Nevertheless, several studies (e.g., FitzPatrick et al., 2011) have indicated that students’ clicker answers are often collected for participation marks rather than right answers. These results suggest the need for a deeper understanding of how the policy of awarding grade points for clicker answers might influence student learning processes and outcomes (White et al., 2011).

### Product factors

Table 4. Description and examples of product factors

<table>
<thead>
<tr>
<th>Product Factors (Frequency)</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Cognitive outcomes (f = 11)</td>
<td>Factors that were used to evaluate the effects of clicker technology(e.g., engagement)</td>
<td>Hogan &amp; Cernusca (2013), Holland, Schwartz-Shea, &amp; Yim (2013), King (2011), Laxman (2011), Oakes &amp; Demaio (2013)</td>
</tr>
<tr>
<td>Cognitive and Non-Cognitive outcomes (f = 28)</td>
<td>These combined factors were used to assess both aspects of cognitive and non-cognitive outcomes.</td>
<td>Johnson &amp; Lillis (2010), Nielsen, Hansen, &amp; Stav (2013), Patterson, Kilpatrick &amp; Woebkenberg (2010), Quinn (2010), Yeh &amp; Tao (2013), Velasco &amp; Çavdar (2013)</td>
</tr>
</tbody>
</table>

### Cognitive outcomes

Cognitive learning outcomes were used to assess and evaluate the effects of clicker technology on student knowledge, learning, and changes in knowledge and learning (see Table 4). Exam scores, quiz scores, overall performance, or a combination thereof was used to examine the effectiveness of clicker use on student learning. Although some studies (e.g., Bright et al., 2013) reported some degree of positive correlation between the use of clickers, perceived learning,
and quiz, exam, or specific test results, other studies indicated that these positive student learning outcomes should not be attributed to the use of clickers (e.g., Anthis, 2011).

**Non-cognitive outcomes**

Non-cognitive outcomes comprise the following two types: Engagement and Attendance. Since there is no clear conceptual agreement across studies on the definition of engagement (Fredricks, Blumenfeld, & Paris, 2004), the current study used the term Engagement to cover student involvement and attention. Attendance (with participation) was also frequently included as a non-cognitive measure in the studies (e.g., Bright et al., 2013). Most studies consistently reported that the use of clicker technology or different pedagogical approaches with clickers had an impact on students’ positive perception of their engagement (Freeman et al., 2011), involvement (Bright et al., 2013), attention (Gachago et al., 2011), participation (Denker, 2013), and attendance (Fortner-Wood et al., 2013). These findings align with those of previous literature reviews (Kay & LeSage, 2009), which reported that the use of clicker technology had an impact on student non-cognitive outcomes, such as an increase in student class attendance and a decrease in course failure rate (Freeman et al., 2011).

**Relationships among presage, process, and product factors**

![Figure 3. Results of heat-mapping (http://mapping-2014.weebly.com).](http://mapping-2014.weebly.com)

The heat-mapping visualization approach was used to identify missing links in the existing clicker studies. The frequencies of all factors and tabulated factors were categorized into Presage-Student and Presage-Instructor factors in the left column and Process-Product factors in the first row in Figure 3. All cross-tabulated information was presented using multi-way contingency tables, which clearly indicated the frequency percentage of co-occurrences of factors in this study. For instance, the results clearly indicated that the relationship between pedagogical training and all other process and product factors was infrequently examined in relation to learning with the use of clickers; thus,
cross-tabulated cells connected to pedagogical training were not colored except in the Assessment and Feedback column. By contrast, the relationship between cognitive learning outcomes and instructional activities/course design factors was frequently examined and thus marked with a large, red rectangle between the cross-tabulated cells in Figure 3. The heat-mapping visualization approach was used to facilitate the comprehensive understanding of the factors examined in this study and the relationships among these factors. The results of heat-mapping indicated that the following relationships were frequently examined in the clicker studies: the relationships between instructional activities/course design as a process and student cognitive outcomes as a product (18.92%); delivery methods and student cognitive outcomes (14.86%); and instructional activities/course design and both cognitive and non-cognitive learning outcomes (9.46%).

Relationships among student presage, process, and product factors

The heat-mapping method (Figure 4) was used to identify the missing relationships among the 3P factors. It should be noted that presage instructor factors were excluded in the further heat-map results because the frequency of instructor presage factors \( f = 9 \) was small and thus may have scattered the presentation focus. The results of heat-mapping highlighted the most frequently and least frequently examined relationships in the clicker studies. Overall, the following three relationships were the most widely investigated in clicker-use studies: the relationships between instructional activities and cognitive outcomes (21.13%), delivery methods and cognitive outcomes (15.49%), and instructional activities and both cognitive and non-cognitive outcomes (11.27%). In other words, nearly 64.8% of the studies did not consider presage student factors, while 35.2% of studies did consider presage factors as variables in their studies. However, students’ non-cognitive outcomes were not considered as either process factors or as student presage factors in the studies.

Relationships between student presage and process factors

The results indicated that almost 41% of studies examined instructional activities/course design without taking into account student presage factors, whereas student background (7.04%) and non-cognitive factors (7.04%) were frequently examined in relation to instructional activities/course design (see Figure 5). Even though over 5% of studies examined more than one student presage factor in connection to process, non-cognitive, and background factors, these combined presage factors were not examined in connection with assessment and feedback.
Relationships between student presage and product factors

Figure 6 highlights the results pertaining to the relationship between presage student and product factors in the clicker-use studies. Specifically, 64.79% of studies examined all three types of learning outcomes without considering student presage factors. Conversely, both cognitive and non-cognitive outcomes were often investigated in the context of all presage student factors (21.13%). Interestingly, no study examined all the presage student factors in relation to non-cognitive outcomes. Cognitive outcomes were likewise not examined in relation to more than one presage student factor in any of the clicker studies.

Relationships between process and product factors

The results indicated that the relationship between process and product factors was broadly examined in clicker studies (Figure 7). Even so, relationships between non-cognitive outcomes and delivery methods, and between non-cognitive outcomes and assessment and feedback, were missing from the studies. In addition, in comparison to delivery methods, assessment and feedback approaches, instructional activities (8.45%) were minimally explored in relation to product factors.
Evidence of learning outcomes: Perceptions and achievement

The heat-mapping method generated a cross-tabulation between three types of learning outcomes and products (Figure 8). Although some studies (22.37%) assessed student cognitive learning based on both student achievement (e.g., quiz score, exam score, performance) and student-perceived learning, nearly 16% of studies measured student cognitive learning using only their perceptions. Student perceptions were also used to assess both cognitive and non-cognitive outcomes (22.68%). Overall, a collection of student perceptions (using survey items and instrument measurement) was the most frequently used method of assessment for student learning in clicker-use studies.

While most clicker studies reported cognitive outcomes as their concrete measure of student learning (see Table 4), many studies (e.g., Fortner-Wood et al., 2013) also investigated student perceptions of learning. Several studies (e.g., Sprague & Dahl, 2010; Schmidt, 2011) identified an interesting and somewhat positive relationship between student perceptions of learning (as measured by survey items such as “I learned a lot of content”) and cognitive learning.
outcomes (e.g., quiz results). Entwistle (2010) provided insight for understanding the nature of the relationship between perceptions and concrete results: he concluded that students’ perceptions of their learning (or learning environments) could be informative when evaluating the degree to which one instructional factor was synchronized with student learning outcomes. Specifically, students’ perceptions might not have a direct effect on student cognitive learning, but were mediated by teaching and learning processes (Trigwell, 2010).

**Methodological issues**

The following five methodological issues and concerns emerged from the literature review. First, most studies on the use of clickers do not report or specify the effect size. Reporting effect size can facilitate the process of interpreting the results and understanding the magnitude or estimated power of the variables in relation to the population (Cohen, 1994). Second, validity and/or reliability test results were not reported in most studies, even when new designs or instruments were developed to measure the effects of learning outcomes. Other reports on clicker studies (e.g., Han & Finkelstein, 2013) have pointed out that a lack of validity and reliability in the scales used may decrease in the credibility of the constructs. Third, although measuring a single factor (such as “satisfaction”) with a single item has met with success (Wanous, Reichers, & Hudy, 1997), it is not clear that measuring a complex construct (e.g., engagement and learning) with a single item can be equally valid. Using multiple items to measure these complex constructs would be preferable. Fourth, some studies that have chosen to explore the effects or mechanism of clickers through qualitative methodologies have failed to fully describe or specify the trustworthiness of the chosen measures (Lincoln, Lynhan, & Guba, 2011). In particular, inter-rater reliability and coding procedures are often not clearly described, an oversight which decreases the credibility of the study results. Fifth, although parametric analysis is the most frequently used method in most clicker studies, normality tests or descriptions about the normal distribution are generally not elaborated upon. Failure to test normality in studies using parametric analysis may threaten the credibility of the results and interpretation.

**Summary**

While other literature reviews on clicker use have provided insights on the advantages and barriers involved in the practical aspects of designing and implementing clicker technology, the current review extends those insights by examining clicker studies in the context of the 3P model, which has been conceptually and practically validated internationally. The current review is also distinct in its use of information visualization methods, including heat mapping, to analyze hierarchical data in the review results. A primary goal of this review is to provide the tools necessary for achieving a better, more integrated understanding of clicker studies. The 3P model and heat-mapping approaches recursively enable us to comprehend the relationships among student and instructor presage factors, process factors (instructional activities), and product factors (outcomes) within the context of clicker use in classroom instruction.

This study showed that the incorporation of guided review protocols from previous systematic literature review studies allows us to pinpoint outstanding methodological concerns related to the rigorous use of research methods in clicker-use studies. This study confirmed the findings of Kay and LeSage (2009) that most studies to date have not reported appropriate information regarding reliability and validity, and thus their results might be anecdotal (Caldwell, 2007). The current review extended these findings by providing evidence that most clicker studies also did not provide effect sizes or description of scale development. Thus, it seems likely that the fault does not lie in a failure to adopt a specific research design, but in a failure to specify and describe the appropriate procedures for enhancing the quality of the evidence.

Specifically, the current review found that recently published studies relied concurrently on perceptions and cognitive/non-cognitive outcomes, rather than either factor in isolation, to measure the effects of clickers on student learning. Past clicker literature reviews (e.g., Kay & LeSage, 2009) have indicated that about a third of the studies reviewed used only learning outcomes as a measure for determining the effect of clickers on student learning; however, the current study found that more than half of studies published after 2009 used more than one measurement method to enhance the validity of their findings. Furthermore, the product factors (cognitive and non-cognitive outcomes) were also simultaneously used as measures to assess clicker effects on learning.
Conclusions and implications

Several studies have adopted distinctive research methods and approaches to explore the impact of clicker use on student learning; the focus of these studies has shifted from the comparison of different methods toward an examination of the relationships among clicker question design, instructional activities, and course design, on the one hand, and various types of learning outcomes, on the other. A relationship between student non-cognitive and/or cognitive learning outcomes and process factors (e.g., clicker question design) was consistently identified in these studies; it may be concluded that more effective process factors enable student to adopt deeper approaches to learning and to actively engage in instructional activities with the use of clickers. All of the findings from the current survey of clicker studies reinforce the message that a balanced understanding of the use of clicker technology in higher education has yet to be achieved. In other words, there are still missing links in our understanding of the relationship between clicker use and education; future studies should address these links in order to enhance our understanding of relationships among the factors.

Missing links and opening the relationship among the factors

Rather than seeking to identify strict cause-effect relationships, the 3P model allows us to focus on interactions between multiple personal, interpersonal, technological and environmental factors related to clicker technology. The use of heat-mapping supports this relational understanding of clicker use and enables us to identify the following three missing links and gaps in the relationships that affect learning:

- One of the most important missing links concerns presage factors related to student learning. Since only a few studies have seriously considered and examined these factors, it is unclear how those factors are synchronized with instructional processes and learning outcomes in the context of clicker use. In other words, the benefits that students derive from clicker use may vary according to their previous achievement levels and their levels of motivation and metacognition (Brady et al., 2013a). It thus is essential to develop a comprehensive understanding of the relationship between the use of clickers and learning.

- Another missing link is the relationship between instructor presage factors and student learning. It is worth noting that instructor-related presage factors have been under-examined in most clicker studies. As a result, the effect of instructor presage factors on student learning processes and outcomes is not yet fully understood, despite the fact that these factors obviously play a key role in the implementation of clicker technology use in the classroom.

- A final missing link to be taken into consideration is the relationship between process and product factors. Specifically, the following process factors might help explain why and how clicker use influences student engagement and learning: types of student interaction with peer groups (Macarthur & Jones, 2013), difficulty level and types of clicker questions (e.g., Cline et al., 2013). Investigating different instructional activities and course designs incorporated with clicker technology may help us to understand what types of instructional strategies show the best-fitting results, and lead to a richer understanding than an investigation that focuses exclusively on the binary effect of use/non-use of clickers on instruction.

Based on the missing links identified above, we suggest that seeking answers to the following research questions will enhance our collective understanding of the use of clickers in higher education: How do the following unexamined factors relate to the use of clickers in classroom learning environments? How do these factors (in conjunction with previously examined ones) influence the process of teaching and learning with clicker technology?

- Student presage factors: individual student goals, interests, and epistemic beliefs
- Instructor presage factors: pedagogical development and course design experience
- Process factors: student (deep/surface) approaches to learning and instructor (content-/student-centered) approaches to teaching
- Product factor: the (coherent or contradictory) relationship between student perceptions and objective evidence of learning
Relationship between the effective use of clickers and the pedagogical issues

Research into the effects of clicker technology tends to have as its goal the identification of implications for the design and implementation of clickers in college classroom instruction. The studies examined in this review have commonly indicated that effective use of clickers is closely related to instructors’ pedagogical view and/or knowledge; thus, the use of clickers may be seen either as “teacher-centered” or “student-centered.” For instance, clickers may be incorporated into active learning strategies to enhance engagement and learning, or they may be used in a more teacher-centered way - for instance, to monitor student attendance rather than to encourage student participation. This issue may be closely related to instructors’ preparedness for clicker use, instructors’ approaches to teaching, and instructors’ pedagogical and technological views and knowledge.

One workaround to resolve the issues stemming from a lack of instructors’ pedagogical and technological knowledge and skills is to provide instructors with the appropriate training and instruction (e.g., Han, 2014). A few studies have suggested that pedagogical training will allow instructors to re-examine their teaching approaches and to enhance their course design knowledge while keeping clicker technology in mind. In terms of specific pedagogical advice, the studies commonly suggest that clickers should be able to collect student logs and data regarding student actions (e.g., participation, answers to the questions); using clickers in this way allows instructors to identify places where student understanding is lagging or to identify misconceptions or naïve conceptions concerning course content. Additional training on clicker data analytics could give instructors the opportunity to re-examine their instructional activities and realign their course design to optimize the course content, goals, and pedagogical approaches with clickers.

Concluding remarks

The current study reviewed the literature regarding the use of clickers in higher education by using the 3P model for teaching and learning. Although research into clicker technology has shown fairly clearly that clicker technology as a pedagogical tool impacts both teaching and learning in classroom instruction, it is still not fully apparent how and why the use of clicker technology affects student learning. Using the 3P model and two visualization methods, the current literature review identified several missing links among presage, process, and product factors and the use of clicker technology. Since teaching and learning with clicker technology is a complex as well as relational phenomenon, a comprehensive understanding of how and why these factors are synchronized or out of sync with other instructional factors will close the missing links and help to fill new gaps that emerge through further studies.

Acknowledgments

This study was conducted with funding support from the Arthur and Dawn Weston Fellowship in Teaching and Learning in Higher Education, McGill International Doctoral Students Awards (MIDAs), and Teaching and Learning Services (TLS), McGill University, Montreal, Canada. The author appreciates constructive comments and suggestions from anonymous reviewers.

References


Murphy, K. (2012). Using a personal response system to map cognitive efficiency and gain insight into a proposed learning progression in preparatory chemistry. *Journal of Chemical Education, 89*(10), 1229-1235. doi: 10.1021/ed100859n


Oakes, C. E., & Demaio, D. N. (2013). “I was able to have a voice without being self-conscious”: Students’ perceptions of audience response systems in the health sciences curriculum. *Journal of Allied Health, 42*(3), e75-e80.


Review of Augmented Paper Systems in Education: An Orchestration Perspective

Luis P. Prieto¹*, Yun Wen¹, Daniela Caballero² and Pierre Dillenbourg¹
¹CHILI Lab, École Polytechnique Fédérale de Lausanne, Switzerland // ²Pontificia Universidad Católica de Chile, Santiago, Chile // luis.prieto@epfl.ch // yun.wen@epfl.ch // dccaball@uc.cl // pierre.dillenbourg@epfl.ch
*Corresponding author

ABSTRACT
Augmented paper has been proposed as a way to integrate more easily ICTs in settings like formal education, where paper has a strong presence. However, despite the multiplicity of educational applications using paper-based computing, their deployment in authentic settings is still marginal. To better understand this gap between research proposals and everyday classroom application, we surveyed the field of augmented paper systems applied to education, using the notion of “classroom orchestration” as a conceptual tool to understand its potential for integration in everyday educational practice. Our review organizes and classifies the affordances of these systems, and reveals that comparatively few studies provide evidence about the learning effects of system usage, or perform evaluations in authentic setting conditions. The analysis of those proposals that have performed authentic evaluations reveals how paper based-systems can accommodate a variety of contextual constraints and pedagogical approaches, but also highlights the need for further longitudinal, in-the-wild studies, and the existence of design tensions that make the conception, implementation and appropriation of this kind of systems still challenging.

Keywords
Paper computing, Augmented paper, Classroom, Orchestration, Review

Introduction

In spite of the high investments made, information and communication technologies (ICTs) are still not fully integrated in the everyday practice of our classrooms (Cuban, 2001), which are still dominated by other legacy tools like pen and paper. Human-computer interaction (HCI) researchers have long proposed to exploit this ubiquity of paper in our everyday life to better integrate computing into it (Wellner, 1993), through paper-based computing technologies, also termed “augmented paper” (Mackay & Fayard, 1999). This broad notion includes not only the use of augmented reality (AR) along with paper objects, but also any technology that uses paper artifacts (e.g., documents) as interfaces to the digital world (Kaplan & Jermann, 2010).

Researchers have proposed to apply these technologies to education, under the rationale that paper, so deeply ingrained in educational practices, would provide seamless interaction with digital educational artifacts (e.g., augmented educational books, digital pen note-taking systems). Over the years, advances in computing power, computer vision and other related technologies have made this kind of systems increasingly affordable, to the point that we can see such products for mass market consumption (such as Sony’s Wonderbook™, http://www.sony.com/pottermore/). However, even though such systems are now affordable for schools, their deployment in authentic educational settings is virtually non-existent. We cannot help but wonder about the reasons for this gap between research proposals and classroom implementation.

This paper attempts to answer that question by performing a systematic review (Kitchenham & Charters, 2007) of the research proposals that apply paper-based computing systems to education, trying to organize and classify their advantages and constraints. In a second stage of analysis, we identify those proposals that have been evaluated in authentic educational contexts, and try to understand what might be the missing link between the variety of research proposals and the lack of actual deployment. To aid us, we use the notion of “classroom orchestration” (Dillenbourg, Järveli, & Fischer, 2009), which is related to the application of novel technologies in the complex multi-constrained setting of an authentic classroom (as opposed to research done in a lab or other controlled settings) (Roschelle, Dimitriadi, & Hoppe, 2013). This paper thus concentrates on the available evidence beyond single-user usability analyses (as suggested by previous reviews of the field, Cheng & Tsai, 2013), to the wider circle of how these (paper-based) technological innovations fit in the socio-technical system of a classroom (Dillenbourg et al., 2011).
Paper-based computing

As noted by Sellen and Harper (2002), paper is still around in many of our everyday activities and settings, despite the undeniable advantages of digital media. Our schools and university classrooms are one of the clearest examples of this resilience of paper: it is very difficult to find one where books, notebooks and other paper elements are not used profusely.

Researchers have argued that this resilience is connected to the way paper has been entangled with our practices over centuries (Johnson, Jellinek, Klotz, Rao, & Card, 1993), due to paper’s unique affordances: we find it easier to read compared to a screen, it is easy to annotate and organize, cheap, can be drawn on, and it is portable (Sellen & Harper, 2002; Kaplan & Jermann, 2010; Steimle, 2012). The idea of exploiting the affordances of paper in connection with the digital world can be traced back to HCI research in the nineties: Wellner (1993) proposed a hybrid digital/paper desktop environment to leverage the advantages of both digital and physical documents. Around the same time, Johnson et al., (1993) proposed to use our familiarity to handle paper as a way to control computers more easily (using classic paper forms and optical character recognition to control computer tasks, thus effectively creating a “paper user interface”).

To “augment” paper, several kinds of technologies are often combined. Input devices include cameras, barcode readers, RFID readers, scanners are used to identify and locate the paper artifacts (e.g., in conjunction with visual markers like barcodes), as well as digital pens (e.g., the Anoto™ technology) that allow the capturing of freeform writing, automatically converting it to a digital equivalent. These systems use the typical range of output devices (screens, projections, sounds and, of course, paper itself through printers). These technologies can be combined in many different ways to implement concrete augmented paper applications, which roughly fit into five basic interface form factors (Steimle, 2012):

- **Augmented Cards and Post-Its**: the paper artifact is treated as a physical token that allows accessing and managing digital resources, which are represented by the (paper) physical objects (e.g., Miura, Sugihara & Kunifuji, 2009).

- **Augmented Books**: the book itself has value independently from the digital resources, although usually includes printed markers to link its contents to additional/complementary media (e.g., Chen & Chao, 2008).

- **Augmented Notebooks**: the notebook, initially empty, synchronizes a paper-based and a digital version of the same resource/contents, allowing free handwriting and sketching (e.g., Lee, Maldonado, & Kim, 2007).

- **Augmented Printed Documents**: often work with a pre-printed document, where users can fill in forms, make annotations or mark parts of the document. These actions are then is translated to a digital counterpart (e.g., Kimbell et al., 2005).

- **Augmented Tables, Flipcharts and Whiteboards**: combine paper-based media with interactive tabletop and/or wall displays, allowing a close integration of paper and digital media (e.g., Do-Lenh, 2012).

These paper-based systems have been applied to a wide variety of fields: augmented cartography in tourism, worksheets in museums, desktop office work, workplace meeting applications, collaborative sketching and prototyping, air traffic control, or entertainment and “edutainment” (Choi, 2009; Huang, Hui, Peylo, & Chatzopoulos, 2013; Shaer & Hornecker, 2010; Steimle, 2012). However, in this paper we are especially interested in the application of these systems to education, given that paper is ubiquitous in our schools already. Aside from the fact that paper artifacts are cheap to produce, several authors have noted the intrinsic qualities of paper for educational settings: paper interfaces can be used to integrate computing in schools more seamlessly than mouse, keyboards and screens (Malmborg, Peterson, & Pettersson, 2007); its tangibility and flexibility allow paper to be easily manipulated, carried around, or passed from one student to another (Horn, AlSulaiman, & Koh, 2013; Luff et al., 2007); as a tangible element, paper can help learning spatial skills (Cheng & Tsai, 2013; O’Malley & Fraser, 2004; Schneider, Jermann, Zufferey, & Dillenbourg, 2011), and it can make the activity workflow visible (Dillenbourg et al., 2011). But, before we analyze educational applications of augmented paper, let us look into the notion of “classroom orchestration” and how it can help us understand the educational use and deployment of these technologies in authentic settings.
Classroom orchestration in educational technology research

We are trying to understand why paper computing solutions have not been widely deployed yet in everyday classroom practice. This gap between the variety and number of research proposals and their implementation in everyday classroom practice is actually a common problem highlighted in educational technology research (e.g., Chan, 2011). Under the label “orchestrating learning,” it has been noted as one of the grand challenges of technology-enhanced learning research (Sutherland & Joubert, 2009).

Dillenbourg et al. (2009) define orchestration as “the process of productively coordinating supportive interventions across multiple learning activities occurring at multiple social levels.” Roschelle et al. (2013) note that orchestration aims at “more meaningful research by acknowledging the complexity and variability of classrooms and the mediating role of the teacher.” Dillenbourg et al.’s (2011) explains orchestration as a change in the focus of attention in the design and study of educational technologies, from individual usability or the study of small-group work, to a wider focus on the usability of educational technologies at a classroom level, including the multiple constraints (time, curriculum, etc.) present in real classrooms. A more complete discussion of the multiple dimensions of this notion can be found elsewhere (Prieto, Holenko-Dlab, Abdulwahed, Gutiérrez, & Balid, 2011; Roschelle et al., 2013; Fischer et al., 2013).

Authors highlight a variety of aspects on what orchestration is and how it should be addressed, with a large amount of overlapping with each other, but also containing marked differences (due to the different research perspectives and educational contexts their authors address). In order to help us analyze existing paper computing systems to explore their orchestration potential, we have tried to find which of these elements are key in defining orchestration, by compiling the most-often appearing elements from previous literature on orchestration. We found eight such elements, some of them marking general agreement among orchestration-related authors, others marking disagreements or tensions that can be resolved in different ways (as noted by Roschelle et al., 2013). The agreements are mostly about:

• **Pragmatism/Constraints.** Complying with the specific contextual constraints of the (authentic) educational setting is of the utmost importance: lesson time limitations, classroom space constraints, assessment requirements, teacher energy levels, curriculum relevance, discipline constraints, etc. (e.g., Dillenbourg, 2013; Nussbaum & Díaz, 2013; Roschelle et al., 2013). Regardless of the potential of an innovation in the lab, it will never be adopted if it does not comply with these constraints.

• **Empowerment/Control/Management.** Orchestration is largely about the logistics of managing the different learning activities taking place in the classroom at different social levels, using different tools (Prieto et al., 2011; Dillenbourg, 2013). Classroom technologies should empower teachers (and students) in this management, e.g., by making it more efficient (automations), or by letting users control them easily (Cuendet, Bonnard, Do-Lenh, & Dillenbourg., 2013).

• **Visibility/Awareness/Monitoring.** The perceptual processes taking place during enactment are crucial, to know what is happening and how student learning progresses (Dillenbourg, 2013; Looi & Song, 2013; Balaam, 2013). Technological systems for the classroom should facilitate such perceptual processes, e.g., by supporting visibility of the learning activities, or their assessment.

• **Flexibility/Adaptation.** Learning situations in authentic settings often require making run-time changes to the original plan of the learning experience (e.g., Cuendet et al., 2013; Looi & Song, 2013; Tchounikine, 2013), to address unexpected events (latecomers, interruptions, etc.), and take advantage of emergent learning opportunities and students’ input (e.g., for debriefing activities). Classroom systems should be flexible enough to accommodate such changes and opportunities.

• **Minimalism.** Given this multiplicity of tasks and processes going on in the classroom, technologies should be simple, providing the most-often-needed functionalities (Balaam, 2013; Dillenbourg, 2013; Looi & Song, 2013).

Authors also point out other elements as being very important for orchestration, even if they do not always agree on how to address them, or how to resolve the tensions they represent:
• **Teacher-centrism and sharing the load.** Most authors acknowledge the crucial role that teachers play in the technology-enhanced classroom (Roschelle et al., 2013). However, different educational contexts and different moments call for different “balances of power” between teachers, students and technology. The central role of the teacher in the classroom can lead to an excessive management burden, calling for systems that allow sharing the orchestration load (e.g., with students, see Sharples, 2013).

• **Designing for preparation, appropriation and enactment.** Certain authors focus orchestration more on the run-time management of the learning situation (Dillenbourg, 2013), while others highlight the activities that happen prior to the classroom enactment itself, including the de-contextualized pedagogical and technological design (scripting), the customization of those designs for a concrete learning setting, as well as other preparation work regarding the appropriation of those elements by teachers and students (Hämäläinen & Vähäsananen, 2011; Kollar & Fischer, 2013). Technological systems should make this preparation and appropriation easy.

• **Multi-level integration and synergy.** Orchestration is said to be about aligning or combining the multiple elements in the classroom to achieve a more effective learning experience, e.g., by combining different technologies, by combining individual, group and class activities to enhance the learning about a subject, or even by combining different learning and pedagogical theories in a sort of “theoretical ecumenism” (Dillenbourg, 2013; Kollar & Fischer, 2013; Looi & Song, 2013). This aspect also includes how the system integrates with existing classroom practices, workflows and contents, and with other technologies being used in the classroom (Cuendet et al., 2013).

From these aspects we can gather an operational definition of “technology for orchestration,” as a technology that is usable and effective within the pragmatic constraints of authentic educational settings, by supporting the perception and management of the learning activities (individual, in groups or at classroom level), and/or being flexible enough to be easily integrated in such management. To better understand the applicability of paper-computing proposals to authentic educational settings, we propose to look at existing paper-based computing systems through these eight orchestration aspects.

**Review methodology**

We performed a systematic literature review following the guidelines by Kitchenham and Charters (2007). The concrete goals of this review were: (a) to identify and organize paper-based computing systems applied to education (to acknowledge their advantages and disadvantages); and (b) to further analyze those efforts that performed evaluations in authentic formal education settings from an orchestration perspective (to assess their potential and the outstanding challenges of their use in authentic practice).

Regarding the need for such a review, several authors have performed reviews in related fields such as AR (Grasset, Dunser, & Billinghurst, 2008; Choi, 2009; Billinghurst & Duenser, 2012; Cheng & Tsai, 2013; Huang et al., 2013; Santos et al., 2014) or tangible interfaces (O’Malley & Fraser, 2004; Shaer & Hornecker, 2010). There also exist reviews on paper-based systems in general (Steimle, 2012). However, there exists no review systematically focusing on the educational uses of paper-based computing systems (Lim and Park, 2011, focus only on the augmented book form factor). Especially, there is no review that looks into the ecological perspective of whether these systems “work” in the context of authentic classrooms.

To find relevant literature sources, we used seven well-known online research databases related to education and technology (IEEExplore, Science Direct, Sciverse Scopus, ISI Web of Science, ACM Digital Library, Springerlink and ERIC). Furthermore, we also queried Google Scholar to ensure the inclusion of “grey literature” (Kitchenham & Charters, 2007). Using the query string: (“paper computing” OR “augmented paper”) AND (education) AND (classroom OR school OR university), we obtained 209 references from the databases, and 634 from Google Scholar. We screened these sources for relevance to paper-based computing and education, eliminating studies that did not involve a concrete system/intervention (e.g., literature reviews). After eliminating duplicates (e.g., sources with similar authors describing the same system in similar settings), and adding expert and reviewer recommendations as well as more recent work by authors and research groups detected as relevant, a total of 40 references were left. These are analyzed in the “Review results” section.
To address the two goals of the review, our analysis followed two stages. First, we synthesized the 40 educational augmented paper systems, clustering their purported advantages and affordances for educational settings. In the second stage, we analyzed the 15 references that included system implementation and evaluation in authentic formal education settings (or that tried to emulate such settings explicitly), using the eight orchestration themes presented in the previous section.

**Review results**

**Stage 1: General overview of paper-based systems in education**

Among the 40 analyzed proposals (see Table 1) we can find examples of every kind of form factor (see “Paper-based computing” section above). For instance, Bayon, Wilson, Stanton, & Boltman (2003) propose KidPad, an environment for primary school students to create and retell stories, using augmented cards with barcodes as placeholders for drawings, audio and other media. Shih, Wang, Chang, Kao, & Hamilton (2007) propose augmented books featuring barcodes which, along with PDAs, facilitate access to additional contents (including also tools for teachers to customize such content, see Figure 1). A form of augmented notebook to support creation/sharing of sketches is proposed by Lee et al. (2007). Hitz and Plattner (2004) propose a generic architecture to augment printed documents to enable students to access additional contents using mobile devices (Figure 2). Finally, Cuendet (2013) proposes Tinkerlamp, an augmented table that uses paper elements to control different simulations in logistics vocational education (see Figure 3).

<table>
<thead>
<tr>
<th>Form factor</th>
<th>Prototype only</th>
<th>Usability test</th>
<th>Experimental / controlled</th>
<th>Authentic setting</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Augmented cards/Post-Its</strong></td>
<td>Ferdinand, Müller, Ritschel, &amp; Wechselberger, 2005</td>
<td>Liarokapis &amp; Anderson, 2010; Montemayor, 2003</td>
<td>Bayon et al., 2003; Kerawalla, Luckin, Seljeftol, &amp; Woolard, 2006; Miura et al., 2009</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td><strong>Augmented books</strong></td>
<td>Ha, Lee, &amp; Woo, 2011; Heger &amp; Lucero, 2008</td>
<td>Grasset et al., 2008; Horn et al., 2013; Wang &amp; Chang, 2007</td>
<td>Chen &amp; Chao, 2008; Martín-Gutiérrez et al., 2010; Shih et al., 2007</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>Augmented notebooks</strong></td>
<td>Pietrzak, Malacria, &amp; Lecolinet, 2010</td>
<td>Lai, Chao, &amp; Chen, 2007</td>
<td>Alvarez, Salavati, Nussbaum, &amp; Milrad, 2013; Lee et al., 2007; Liao, Guimbretière, &amp; Anderson, 2007; Mitsuhara, Yano, &amp; Moriyama, 2010; Steinle, Briczka, &amp; Muhlhauser, 2009</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td><strong>Augmented printed documents</strong></td>
<td>Hitz &amp; Plattner, 2004; Signer &amp; Norrie, 2007</td>
<td>Fraser et al., 2003; Luff, Pitsch, Heath, &amp; Wood, 2010</td>
<td>Alessandrini, Cappelletti, &amp; Zancanar, 2014; Chen &amp; Tsai, 2013; Lu, 2008; Teng, Chen, &amp; Lee, 2011</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
Paper-based systems have been proposed for almost every educational level: from preschool and elementary school storytelling environments (e.g., Bayon et al., 2003), to secondary education books (e.g., Wang & Chang, 2007), vocational training tabletop systems (e.g., Do-Lenh, 2012) and university-level lecturing support systems (e.g., Mitsuhara et al., 2010). Similarly, augmented paper systems have been proposed to aid in a wide variety of subject
matters, including math (e.g., Sharma et al., 2013), science (Kerawalla et al., 2006), history (e.g., Fraser et al., 2003), etc. Other efforts focus on specific skills rather than subject content: storytelling (Bayon et al., 2003), spatial skills (e.g., Martin-Gutierrez et al., 2010) or even guitar playing (Liarokapis & Anderson, 2010). Yet others are aimed at supporting activities more peripheral to learning, such as note-taking (e.g., Pietrzak et al., 2010), and thus could be applied to learning multiple kinds of content/skills. The proposed systems cover learning activities throughout the whole spectrum of social planes: from individual activities (e.g., Chen & Chao, 2008) to small group activities (e.g., Hook et al., 2013) or even whole-class collaborative learning activities (e.g., Alvarez et al., 2013).

Table 2. Synthesis of augmented paper advantages for education, including examples

<table>
<thead>
<tr>
<th>Role of paper</th>
<th>Paper documents contain hyperlinks (e.g., visual markers)</th>
<th>Contents of paper are synchronized with digital resources</th>
<th>Paper as a tangible token to access/control digital content</th>
<th>Hybrid of augmented document and tangible controller</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>From paper’s intrinsic properties</td>
<td>From digital resources’ intrinsic properties</td>
<td>From mutual advantages of paper and digital</td>
<td></td>
</tr>
<tr>
<td>Flexibility</td>
<td>Augmented books [Ha et al., 2011; Zhao et al., 2014]</td>
<td>Augmented printed documents [Hong &amp; Jung, 2005]</td>
<td>Augmented paper cards with digital pens [Miura et al., 2009]</td>
<td></td>
</tr>
<tr>
<td>Enables complex interaction preventing cognitive/visual overload</td>
<td>Augmented printed documents [Lai et al., 2007; Lee et al., 2007; Mitsuhara et al., 2010; Steimle et al., 2009]</td>
<td>Augmented printed documents [Lu, 2008]</td>
<td>Augmented printed documents [Lu, 2008]</td>
<td></td>
</tr>
</tbody>
</table>
If we look at the system rationales and advantages portrayed by their authors, we can synthesize several augmented paper affordances for education (see Table 2 for more details and example references). Some of these advantages directly stem from paper’s intrinsic properties. Paper can be easily and intuitively used, and thus can provide a means for interaction with less perceived effort, e.g., the ease of navigating through content by flipping pages on an augmented book (tangibility or manoeuvrability). Paper can also support a range of activities (e.g., passing it around to others, or moving it in the area of a desk), thus making it simple to configure digital systems by moving paper around (flexibility).

Other systems leverage digital resources’ intrinsic advantages. Since digital resources can be linked or animated at a distance, and they can be easily searched, shared, duplicated or archived, these systems can enable an immersive experience including visual, audio and other media (e.g., enhancement of a paper map with multimedia content). They also can make the learning process and outcomes more explicit and accessible, through recording and tracking interactions with the systems (e.g., recording of the sequence of logistic simulation parameters tried out by a group of students).

Authors also mention other advantages of these systems, which relate to the benefits of combining paper and digital media. The expansion of the interaction space resulting from linking paper documents and digital resources enables complex interaction without excessive cognitive overload. These systems also bridge the gap between physical and virtual objects, which may assist in enhancing students’ visual and spatial abilities (e.g., projecting abstract notions like angle measures over a paper polygon). Also, the seamless transition between paper and digital documents enables effective information searching, quick navigation, or convenient conversion, storage and retrieval.

However, what kind of evidence do we have of all these advantages? Many of the analyzed proposals (47.5%, see Table 1) only provided a first prototypes of the system, or preliminary pilots or user studies (and thus there is little evidence of their affordances in actual educational use). Other studies perform controlled experimental designs, often in a lab setting (15%), and thus their ecological validity in an authentic, non-controlled setting is limited.

In fact, only 15 of the selected studies (37.5%) depict the usage of the system in an authentic setting, or try to mimic real classroom constraints (in terms of space, time, curriculum, etc.). We could ask ourselves to which extent those affordances are useful for teachers and students within the multiple constraints of actual classrooms.

Stage 2: Orchestration analysis of paper-based classroom interventions

From the 15 studies that had been evaluated in authentic or authentic-like settings (see Table 3), we can also extract several features. Most of them are relatively recent, maybe indicating that only lately the involved technologies have been reliable and affordable enough to be studied out of a controlled lab setting. The studies use a variety of research designs and methodologies, from quasi-experimental designs with a control group to design-based research and participatory longitudinal studies. Although the studies report a wide range of findings, there is a notable lack of studies actually studying learning effects of using the technology, with many of them rather focusing on other constructs like usability (Bayon et al., 2003) or student engagement (Alvarez et al., 2013; Miura et al., 2009). From those studies that try to measure learning effects, many do not find statistically significant results when compared with learning the same material using other methods (e.g., Cuendet, 2013).

Among these 15 systems there are also representatives of all five interface form factors. The proposals provide different levels of learning activity support: some focus especially on the provision of subject matter content (e.g., the 3D models in Martín-Gutiérrez’s augmented book, 2010), others on the content-independent support of a specific task (e.g., Bayon et al.’s environment for storytelling in which teacher and students create freely the contents to be added, 2003), and yet others focus on supporting a learning activity as a whole (e.g., Cuendet’s tabletop system provides both contents and task support to enhance carpenters’ spatial skills, 2013). Looking at these 15 studies from the point of view of the eight key orchestration elements (see the “Classroom orchestration” section above), we notice the following:
Pragmatism/constraints

The analyzed efforts comply with many of the constraints of the real classrooms they were designed for. Most of the studies used curriculum-relevant contents and activities, often co-designed with teachers (e.g., Bonnard’s geometry activities based on the Swiss primary school curriculum, 2012). They also considered classroom space constraints (e.g., using webcams and displays already existing in the classroom, in Kerawalla et al., 2006) and time constraints (e.g., lesson length of a typical lecture in Miura et al., 2009). The impact of the system/intervention into teachers’ energy levels is less clearly addressed in general: while some studies appear to have no impact on teacher effort (e.g., Lee et al.’s student note-taking system, 2007), or even appear to require less effort than the traditional enactment alternative (e.g., Alvarez et al.’s automation of the data flow between digital pens and shared display, 2013), several others mention a certain teacher effort involved in the preparation of materials (Mitsuhara et al., 2010), or the improvisation of debriefing activities (Chen & Chao, 2008). Interestingly, several studies mentioned how the limited range of applicability of the system (e.g., the fact that their usage covers only a small part of the curriculum) may affect adversely system adoption (Bonnard, 2012; Kerawalla et al., 2006).

Table 3. Summary of the augmented paper systems used in authentic settings

<table>
<thead>
<tr>
<th>Form factor</th>
<th>Study</th>
<th>Aim</th>
<th>Subject</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented paper cards with digital pens</td>
<td>Bayon et al., 2003</td>
<td>Enhance children’s creativity and collaboration</td>
<td>Storytelling</td>
<td>* It was easier to draw with pen/paper, rather than with the mouse</td>
</tr>
<tr>
<td></td>
<td>Kerawalla et al., 2006</td>
<td>Incorporate AR content into UK primary school lessons</td>
<td>Earth sciences</td>
<td>* Child participation and engagement lower in AR role-plays</td>
</tr>
<tr>
<td></td>
<td>Miura et al., 2009</td>
<td>Facilitation of collaborative and interactive learning in regular lectures</td>
<td>(Generic)</td>
<td>* Improved student motivation/enjoyment (enjoyable competition)</td>
</tr>
<tr>
<td>Augmented books</td>
<td>Shih et al., 2007</td>
<td>Compare ubiquitous, multimodal e-Learning with alternatives</td>
<td>(Generic)</td>
<td>* Increased information overload when using augmented books</td>
</tr>
<tr>
<td></td>
<td>Chen &amp; Chao, 2008</td>
<td>Explore context-aware learning support to assist in paper textbook comprehension</td>
<td>Programming</td>
<td>* Learning disorientation turned up easily when reading online</td>
</tr>
<tr>
<td></td>
<td>Martin-Gutiérrez et al., 2010</td>
<td>To investigate the use of AR for improving spatial abilities of engineering students.</td>
<td>Engineering</td>
<td>* AR application was helpful for improving student spatial abilities</td>
</tr>
<tr>
<td>Augmented paper notebooks</td>
<td>Alvarez et al., 2013</td>
<td>Support New Media Literacies teaching and curriculum in classroom</td>
<td>Math</td>
<td>* The system can be well integrated in classroom teaching</td>
</tr>
<tr>
<td></td>
<td>Lee et al., 2007</td>
<td>Ecologically valid paper-based system and effects on design culture</td>
<td>Design education</td>
<td>* Longitudinal impact of augmented paper interactions on design practice</td>
</tr>
<tr>
<td></td>
<td>Liao et al., 2007</td>
<td>Combine physical artifacts (paper) with communication and archival infrastructure</td>
<td>Engineering</td>
<td>* Team dynamic effect upon technology appropriation</td>
</tr>
<tr>
<td></td>
<td>Mitsuhara et al., 2010</td>
<td>Compare learning effects and note-taking behavior of augmented paper</td>
<td>(Generic)</td>
<td>* Changes in note-taking behavior: less changes in gaze direction, less time spent writing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>* Little difference in learning effects observed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>* Weak inference or statistical power over student parameters like</td>
</tr>
</tbody>
</table>
Empowerment/control/management

Several issues come to our attention when we analyze the management of classroom activities using paper-based systems. In many of the analyzed systems the teacher is the central “manager” of the flow of the class. However, only a few of the systems provided teacher-specific user interfaces with special capabilities (e.g., Do-Lenh, 2012; Steimle et al., 2009). Certain proposals are specifically designed to facilitate the management of emergent activities like debriefing (Kimbell et al., 2005). Regarding the mode of activity management (socially vs. technology-mediated), very often the proposed systems are managed flexibly in a social manner, without intervention from the system (e.g., Alvarez et al.’s social regulation of the use of the digital pens by students when needed, 2013). Several systems were conceived for easy/flexible navigation, without a specific activity flow (e.g., Lee et al., 2007), although an order is sometimes suggested by the paper binding of a book (as in Martin-Gutierrez et al., 2010). In fact, certain systems did use the paper elements to embody the class workflow (e.g., in Bonnard’s sheets printed with the activity description - see Figure 4), which can even be taken out of the classroom to interact with external actors (Do-Lenh, 2012).

![Figure 4: Bonnard’s (2012) printed sheets including activity description](image-url)
Visibility/awareness/monitoring

Several levels of awareness are possible in the classroom, related to the social plane at which the learning activities happen (individual, small-group or classroom-wide). Digital pen systems often provide very limited visibility beyond the individual; tabletop systems such as Bonnard (2012) normally have good individual or small-group visibility, but not so good classroom awareness from a distance. Thus, there are systems that complement the paper-based system with a shared display in the front of the classroom, as a collective memory and to help teachers be aware of student actions (e.g., Alvarez et al., 2013; Miura et al., 2009). Some authors note that the physical layout of the system elements (e.g., the paper pieces) in the class may help signal aspects of classroom workflow (e.g., Cuendet, 2013). Aside from shared displays, a few systems include specific tools for teacher awareness (e.g., to follow the navigation patterns and learning outcomes of possibly distant students, in Shih et al., 2007). Several systems also provide automated feedback about the learning task to students/teachers (e.g., Chen & Chao, 2008; Cuendet, 2013).

Flexibility/adaptation

A paper-based system can be flexible and adaptable in many different ways. Many of the portrayed systems are flexible to navigate, thanks to paper’s handling ease (e.g., using barcodes on paper cards to freely navigate between a story’s events, in Bayon et al., 2003). However, the fact that some systems provide multimedia learning content poses a limitation to this flexibility, as the lesson cannot step out of the previously prepared contents (as in Mitsuhara et al., 2010). Conversely, systems that are content-independent (e.g., Alvarez et al., 2013; Steimle et al., 2009) may have a larger range of applicability. Other systems use paper elements like cards to flexibly adjust system features and the difficulty of learning tasks (Cuendet, 2013), and the automatic recording of student information and learning outcomes can be used for spontaneous debriefing activities (e.g., Miura et al., 2009).

Minimalism

The reviewed systems comply with this guideline to different degrees. To obtain a simple activity flow, certain systems propose simple, modular activities without many inter-dependencies among them (e.g., Bonnard, 2012). Several systems try to minimize their feature set by doing away with the notion of login/identification, which often are not essential for learning itself (e.g., Do-Lenh 2012; Kerawalla et al., 2006). Strategies to minimize orchestration load include the automation of the workflow (e.g., transparent/automatic sharing of student outcomes in Lee et al., 2007). Yet, we also find systems breaking this simplicity principle, either increasing user perception of cognitive load (e.g., Shih et al., 2007), veering towards clutter due to the multiple tangible interface elements (noted by Cuendet, 2013) and shared student artifacts (Liao et al., 2007). Certain works also mention teachers’ rejection of complex technology setups (e.g., upon Cuendet’s suggestion of using multiple augmented tabletops in a classroom).

Teacher-centrism and sharing the load

Paper-based systems can support very varied role distributions among actors, depending on the classroom dynamic to be supported. In a teacher-centric situation, several systems support “traditional” roles such as teachers lecturing with students as passive receptors (as in Mitsuhara et al., 2010), although more often teacher is seen as a facilitator (especially in creative activities, like storytelling in Bayon et al., 2003). On the other end, certain systems are designed to work independently of the presence of a teacher, for student individual study or problem resolution (e.g., Martín-Gutiérrez et al., 2010; Shih et al., 2007). Some of the systems actually provide more flexible orchestration roles, enabling teachers to share with students a part of the orchestration “power,” for example in the form of cards enabling certain features (e.g., Bonnard, 2012; Cuendet, 2013). We observe also a great variety of balances between teachers and researchers, especially regarding who (if anyone) has to handle the preparation of the learning activity and its materials (see also the following aspect).
Designing for preparation, appropriation and enactment

Although the design and implementation of all the proposed systems was done by researchers/specialists, the customization and preparation of the learning activities for the concrete classroom took different forms (although it is often insufficiently described). Certain systems require very little physical preparation, especially those that rely more on student input like note-taking (Alvarez et al., 2013; Steimle et al., 2009) than in the provision of multimedia content or feedback. Other systems, however, require a researcher or specialist to intervene during preparation (e.g., of augmented slides before the lectures in Liao et al., 2007). How this preparation is performed varies greatly, and normally no authoring tool is provided (with the exception of the one depicted in Shih et al., 2007). Only a few systems allow for the rapid and flexible preparation of activities just before the lesson (as in Bonnard’s symmetry exercises, 2012), which can lead to easier appropriation of the system by teachers.

Multi-level integration and synergy

Generally, the proposed systems are well integrated with existing classroom workflows, practices and tools (especially paper-based ones, but not only): using paper as a free drawing tool (e.g., Lee et al., 2007), or as a permanent memory (e.g., use of activity sheets, Do-Lenh, 2012). Other systems infiltrate the use of existing elements like rulers and other drawing tools (Bonnard, 2012, Cuendet, 2013 – see Figure 5). Another common tool synergy involves paper-based elements (for more natural individual/group work) and a shared display (for classroom awareness), as in Do-Lenh (2012) or Alvarez et al. (2013). Other systems try to integrate ubiquitously with existing student practices (such as mobile device-based support for individual study, Chen & Chao, 2008). Many of the proposed systems support some kind of transition between the individual, group and classroom social planes (e.g., in Alvarez et al., 2013, the activity flow includes collaborative problem solving at all three levels). However, the interoperability of augmented paper data and activities with other technological tools and systems (e.g., online LMSs) is much less discussed, aside from the physical persistence of paper artifacts themselves.

Discussion: Challenges and future research for augmented paper systems in the classroom

As we have seen, researchers have proposed augmented paper (in its many forms) as more subtle way to enhance the classroom with digital media. From the first stage of our systematic review, we elicited several potential affordances of this kind of systems, like tangibility or manoeuvrability, the ability to manage documents in an effective way, or to bridge the gap between physical and virtual worlds. We also saw how paper-based computing applications for education is still a field in flux, exploring creatively divergent lines of research and system form factors, but with little continuity beyond preliminary usability/engagement studies with a few subjects. Strangely enough, the actual affordances for learning itself often are not thoroughly investigated, and comparatively few studies address teachers’ perceptions and system compliance with real classroom constraints. As noted by Cheng and Tsai (2012) and Santos et al. (2014), despite initial evidence of benefits for learning of spatial skills or science conceptual understanding, the number of studies is still small, especially for longer, in-the-wild measurements of learning (an issue also highlighted in Lee et al., 2007; Shaer & Hornecker, 2010).
In this paper we have focused specifically on this dimension of usage and adoption in authentic educational settings, through the analytical lens of classroom orchestration. Our orchestration-based analysis of the systems evaluated in authentic settings was limited by the lack of descriptions or explicit consideration of certain orchestration aspects. In latest works, however, classroom usage and the learning value of the technology have begun to be taken into consideration (e.g., Alvarez et al., 2013, Do-Lenh, 2012, and others). Nevertheless, the available studies confirm paper’s protean qualities for classroom use, accommodating a variety of classroom dynamics and multi-level activity integration. However, as noted before, the learning effects of the proposed systems often have not been thoroughly studied yet, with clear benefits present only in certain tasks with an important spatial component.

As technology designers, we could take an alternative approach and focus on the orchestration advantages of paper-based systems at the classroom level, rather than on the individual learning effects (although, of course, studies should ensure that learning is not hampered by using the system in a classroom). In this sense, paper-based systems have also shown to be able to accommodate a variety of pragmatic classroom context constraints. The technology design of the augmented paper systems leveraged different paper-based affordances and advantages: intuitive interaction, accommodating emergent debriefings and flexible navigation sequence, as well as the awareness properties of the different system elements at group and classroom social levels (e.g., paper elements as visible workflow, often in synergy with a shared display for further classroom awareness).

From an educational technology designer perspective, we should note that this compliance with the classroom constraints was the consequence of long and careful co-design processes. Paper-based technologies are still difficult to design and prototype, and it is even more difficult to design systems that let end-users (e.g. teachers, students) appropriate them. As Santos et al. (2014) mention, some systems are starting to provide this kind of end-user support. We foresee that the advances in web technologies and mobile devices can make this kind of features possible, as already hinted by mobile AR browsers (see Muñoz-Cristóbal et al. for an example of a system that applies such technologies to let teachers and students shape their learning activities, 2013) and online paper/tangible solutions (e.g., Cuendet, 2013).

Our review also uncovered certain remaining challenges in this field, which represent interesting directions for future research in this area: a) the design space defined by augmented paper advantages and system form factors (see Table 2) is still sparsely populated, indicating potential paths for the design of novel systems of this kind; b) teachers’ understandable resistance to more complex technical setups, the increased length of the co-design processes in order to address these real classroom constraints, and the limited range of applicability of many of the resulting systems point towards the need of specific design guidelines and processes for the conceptualization and implementation of this kind of systems; c) addressing the design, preparation and customization of the learning activities before enactment itself is also unresolved, as it often requires quite specialized technical knowledge and careful interaction design – augmented paper toolkits and user-created paper UIs are another promising avenue of future work.

Finally, the analyzed systems also illustrate certain design tensions (Tatar, 2007) in applying paper-based systems to the classroom: a) flexible, easy to use paper UIs can quickly become scattered (with interface clutter breaking the minimalism principle); b) designing the system as a scarce, unique resource in the classroom, versus having one-to-one setups, in which increased access to technology has to be balanced with minimalism and teachers’ natural fear of complex setups that can break down or become uncontrollable; c) paper’s natural affordance for flexibly navigating through content is counterbalanced by the need of such content to be prepared by teachers or specialized staff; d) the need to support awareness at different social levels, often using an ecology of different devices, again is in tension with the preference for a minimalist setup; e) computers’ ability to provide automated feedback is somewhat hampered if we make use of paper’s power as a flexible input source (due to the difficulty in converting freeform drawing to computer-interpretable form). These and other design tensions also mark pathways to be explored in future research work (e.g., through comparative studies in different classroom settings. Paper is making a comeback to the classroom, only it never really went away.

Acknowledgements

The work presented in this article has been partially funded by the Marie Curie Fellowship MIOCTI (FP7-PEOPLE-2012-IEF project no. 327384).
References


The Effects of Meta-Cognitive Instruction on Students’ Reading Comprehension in Computerized Reading Contexts: A Quantitative Meta-Analysis

Yi-Chin Lan¹, Yu-Ling Lo² and Ying-Shao Hsu³*

¹Department of Early Childhood Education, University of Taipei, Taipei City, Taiwan // ²Department of Special Education, The University of Texas at Austin, USA // ³Graduate Institute of Science Education, National Taiwan Normal University, Taipei City, Taiwan // lollipop0913@gmail.com // jssabrina@gmail.com // yshsu@ntnu.edu.tw

ABSTRACT

Comprehension is the essence of reading. Finding appropriate and effective reading strategies to support students’ reading comprehension has always been a critical issue for educators. This article presents findings from a meta-analysis of 17 studies of metacognitive strategy instruction on students’ reading comprehension in computerized reading contexts. Overall, some instances of metacognitive strategy instruction tended to be more effective than others. Additionally, the effects of the instruction seemed to vary according to participants’ characteristics. Drawing upon the findings of this meta-analysis, we propose recommendations for future research and practice.

Keywords

Meta-analysis, Metacognition, Self-regulation, Reading comprehension, Computerized reading contexts

Introduction

A universal agreement among educators is that the ultimate goal of reading is to comprehend text. As the National Institute of Child Health and Human Development (2000) directly pointed out, “Reading comprehension has come to be the essence of reading” (p. 4-1). Without comprehension, reading is reduced to a mechanistic and meaningless skill (Oberholzer, 2005, p. 22).

The acquisition of reading comprehension skills is critical for every student to gain important information from written texts. However, comprehension is a complex skill; in addition, the skills needed to comprehend texts vary by text form, genre, reader capacity, readers’ prior knowledge, and reading goals (The RAND Reading Study Group, 2002). Thus, acquiring and applying reading strategies is thus in turn critical for students.

Comprehension is facilitated when readers use strategies (Rupley et al., 2009). While some findings have shown that good readers might adopt several effective reading comprehension strategies when reading text or during reading tasks, researchers such as Yuill and Joscelyne (1988) argued that less skilled readers especially benefit from instruction. Thus, it is assumed that learning outcomes in reading are related to the quality of the instruction students receive. Various examples of effective reading instruction and their characteristics, challenges, nature, and other significant features have been mentioned and documented. For example, Rupley et al. (2009) found that new material could be bridged with prior knowledge through an explicitly-instructed, detailed process which includes guided practice.

Among these efforts, metacognition has been identified as a significant factor for text comprehension (e.g., Williams & Atkins, 2009). As Harris’s (1990) groundbreaking conclusion suggested, metacognitive abilities seem to be a differentiating factor between good and poor readers. Therefore, Harris (1990) argued, “there would appear to be some value in teaching students to apply metacognitive strategies” (p. 34). In line with this advocacy, many researchers have been devoted to examining the role of metacognitive strategy instruction in reading comprehension (e.g., Cubukcu, 2008; Dabarera, Renanda, & Zhang, 2014). Yet, only a few particularly aimed at examining how metacognitive strategies helped students comprehend digital texts.

In a world in which electronic reading is becoming increasingly common, the reading platform today has shifted from traditional text to hypertext. Prensky (2001) termed students today as “digital natives.” In his words, they are those who have grown up in a world where technology is ubiquitous. Their libraries are on their laptops and other handheld electronic devices; they typically read electronic books rather than printed books. According to
Puntambekar and Stylianou (2005), this kind of text is nonlinear and flexible, thus requiring learners to engage in cognitive monitoring. When reading such texts, they need to “plan what to read next, and closely monitor ongoing learning” (Puntambekar & Stylianou, 2005, p. 454). In other words, these digital natives “process information fundamentally differently from their predecessors” (Prensky, 2001, p. 1). It then becomes clear that the argument “the same methods that worked for the teachers when they were students will work for their students now” is no longer valid (Prensky, 2001, p. 3).

Nevertheless, Srivastava and Gray (2012) pointed out that little consideration has been given to how the aforementioned shift would help or hinder students’ reading comprehension. Therefore, there is an urgent need to conduct such a study to offer in-service teachers who work with their “digital native” students some practical ways of implementing the most appropriate reading comprehension strategies. This article addresses this necessity by presenting findings from a meta-analysis of published refereed quantitative studies that examine the effects of metacognitive strategy instruction on students’ reading comprehension in computerized reading contexts. The findings from this investigation offer a credible source of information about which form of metacognitive strategy instruction would be more effective in terms of helping students comprehend non-traditional texts. This work also offers a rich source of information on the relationships between students’ important individual characteristics (e.g., language status) and the effect of different metacognitive strategy instruction.

This meta-analysis was designed to answer the following four questions:

- What metacognitive strategies in computerized reading contexts have been investigated?
- What is the effect of metacognitive strategy instruction on students’ reading comprehension in terms of grade level, reading ability level, and their language status?
- What is the effect of metacognitive strategy instruction on students’ reading comprehension in terms of genre of instructional content?
- What is the combination of the effect of metacognitive strategy instruction and the type of computerized reading contexts on students’ reading comprehension?

**Method**

**Data collection**

According to Petitti (2000), meta-analysis follows several specific steps. First of all, studies of a topic need to be systematically identified, then filtered according to the inclusion criteria.

**Article search procedure**

The procedure implemented was a three-step, comprehensive search strategy. First, online databases such as the PsycINFO and Google Scholar were scanned for potential studies.

Descriptors entered into the search engines were combinations of reading comprehension AND metacognition, metacognitive, self-regulated, self-monitoring AND computerized/computer-based/computer-assisted, digital, electronic, multimedia/hypermedia, hypertext, on-line literacy, e-book/e-reading, technology/technological, on-line/web-based, interactive learning environment, new literacy, pad, and Internet. As a side note, the reason that we used metacognition and self-regulated/self-monitoring simultaneously was that these two constructs usually connect with each other or are used interchangeably (e.g., Dinsmore, Alexander, & Loughlin, 2008). Also, given that computerized reading contexts might present in a variety of forms, we used as many terms as possible to refer to this type of environment.

The total number of articles found, after combining the three groups of descriptors, was 57.

The second step was an ancestral search. We examined the references in the 57 articles to ensure that no studies related to our topic were left out. Lastly, in addition to the first two steps, we conducted a thorough manual search of five pioneering journals from 1979 to December 2013: Computers and Education, Computers in Human Behavior, Interactive Learning Environments, the Journal of Literacy Research, and Metacognition and Learning. These
journals were specifically examined because of their prominent role in such fields as metacognition, reading/literacy, and digital learning.

Inclusion and exclusion criteria

The following inclusion and exclusion criteria were set to guide the article selection:

- Studies were published between 1979 and 2013. In 1979, Flavell published his classic work *Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry*, in which he detailed the notion of this construct. Since then, metacognition has been studied in various content areas (e.g., science education). As such, searching studies from 1979 forward offers a complete analysis of metacognitive strategy instruction provided to students of various backgrounds.

- Studies were published in refereed journals in English.

- Studies involved explicit metacognitive instruction. Clearly defining metacognition has been a challenging task (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Therefore, as Zohar and Barzilai (2013) suggested, “to state clearly which of the existing multiple theoretical perspectives they adopt and to explain the meaning they designate to the construct of metacognition and to the metacognitive sub-components they use in their work” (p. 2) would be helpful for communicating with others about how we define “metacognitive strategy instruction” in this paper. We loosely followed Zohar and Barzilai’s (2013) coding scheme, and only used those studies which met any of the nine practices in their framework. Studies that did not involve any form of metacognitive instruction in Zohar and Barzilai (2013) were therefore excluded (e.g., Srivastava & Gray, 2012).

- Studies that did not use any form of electronic text were excluded (e.g., Houtveen & van de Grift, 2007), as were those that did not include any reading comprehension measures (e.g., Greene et al., 2010).

- Studies that adopted an experimental or quasi-experimental design.

- Studies that provided sufficient statistical or quantitative information to allow calculation of the effect size.

After finalizing the inclusion criteria, we then went back to review the titles and abstracts of the potential 57 articles to determine whether they qualified for further analysis. Those studies which included metacognitive strategy instruction aiming to improve the reading comprehension of electronic texts, and outcome measures which included at least one reading comprehension measure were selected from this potential pool. As a result of this data selection process, the final data set included 17 studies reported in 14 articles (see Appendix) that were peer-reviewed, based on original research, took place in a variety of computerized reading contexts, and involved students of diverse backgrounds. This size was considered acceptable since a) there is no specific rule for minimum number of studies, and b) prominent researchers in the meta-analysis area have used less than 20 studies before (e.g., Mol et al., 2008).

Data analysis

Coding procedures

An extensive coding sheet was adapted from Bryant (2007) for the purpose of the study. Information from the articles was coded in the coding sheet with the following: (a) underlying concept of the study (i.e., purpose of the research, research rationale, research questions), (b) participant description (e.g., age), (c) methodology (e.g., dependent variable(s), (d) intervention condition, (e) control/comparison condition, and (f) findings. Articles were coded by the first two authors using the coding sheet. The information was then summarized into two tables (see Tables 1 & 2).
Reliability

Once the first two authors coded all 17 studies, 4 (24%) were randomly picked to be coded by the other person. Both coders had more than five years of experience teaching and researching in the field of education. Agreement was counted when both coders had the same or similar information for an item on the coding sheet. The mean agreement was 92.15%, ranging from 91.4% to 94%. Through several panel meetings, the disagreements were discussed and resolved for further analysis.

Effect size calculation

Many calculations of effect size (e.g., Cohen’s d, Hedges’ g) can be used to detect the magnitude of effectiveness (Cohen et al., 2011). In this meta-analysis, we used effect sizes to detect the magnitude and strength of the effectiveness of the metacognitive strategies via the computerized reading context.

Cohen’s d is the standardized difference between two populations. For studies where effect sizes were reported, they were recorded directly from the study (e.g., Dreyer & Nel, 2003). For studies where effect sizes were not reported, Cohen’s d was calculated using information provided from the studies. To calculate Cohen’s d, we took means and standard deviations from the studies, subtracted the control group’s mean from the experimental group and divided it by the pooled standard deviation (e.g., Graesser et al., 2007 Study 1). If the mean and the standard deviation were not available, effect sizes were calculated using the F- or t-test scores provided (e.g., Hathorn & Rawson, 2012, Studies 1 & 2). The magnitude of the effectiveness was decided according to Cohen’s (1988) criteria: d = .80 is a large effect, d = .50 is a moderate effect, and d = .20 is considered a small effect.

Results

To lay out the ground for discussion, we first present the study characteristics of the included studies. Then, the types of metacognitive strategy instruction reported in the studies are presented along with the effects. Effects are also presented by participant characteristics and genre. Finally, we analyze the effect of metacognitive instruction and type of computerized reading context on participants’ reading comprehension for the three most and least effective studies, respectively.

Study characteristics

As mentioned, the search yielded 17 studies reported in 14 articles (see Table 1). Of the 14 articles, only two (14%) were published before 2000, while up to 71% were published between 2005 and 2012. These numbers might reveal that interest in examining the effects of such strategies on improving students’ comprehension of computer-based texts has just begun.

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Duration/ frequency</th>
<th>Experimental group condition</th>
<th>Control group condition</th>
<th>Type of computerized reading context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azevedo et al (2007)</td>
<td>82</td>
<td>40 minutes/ once</td>
<td>ERL: Tutor assistance to scaffold self-regulated learning</td>
<td>SRL: No tutor assistance</td>
<td>Hypermedia</td>
</tr>
<tr>
<td>Dalton et al. (2011)</td>
<td>106</td>
<td>24 sessions/ Twice a week</td>
<td>ICON 1: Comprehension strategy ICON 2: Vocabulary ICON3: 1 +2</td>
<td>No control group</td>
<td>ICON SDR environment,</td>
</tr>
<tr>
<td>Study</td>
<td>Duration</td>
<td>Strategy</td>
<td>Comprehension aids</td>
<td>Comparison Tools</td>
<td>Notes</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------</td>
<td>-----------------------------------</td>
<td>--------------------</td>
<td>---------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Dreyer &amp; Nel (2003)</td>
<td>13 weeks/NR</td>
<td>Strategic reading instruction</td>
<td>No strategy instruction, no access to Varsite. No aids</td>
<td>Varsite: a Learning Content Management System Online computer program written in Flash</td>
<td></td>
</tr>
<tr>
<td>Gegner et al. (2009) study 1</td>
<td>Approximately 125 minutes/ Once</td>
<td>Comprehension aids</td>
<td>No aids</td>
<td>Online computer program written in Flash</td>
<td></td>
</tr>
<tr>
<td>Gegner et al. (2009) study 2</td>
<td>Approximately 125 minutes/ Once</td>
<td>Comprehension aids</td>
<td>No aids</td>
<td>Online computer program written in Flash</td>
<td></td>
</tr>
<tr>
<td>Graesser et al. (2007) study 1</td>
<td>50 minutes/ Once</td>
<td>SEEK web tutor: critical stance, inquiry, and self-regulated learning</td>
<td>Navigation only</td>
<td>Google™ search engine and websites</td>
<td></td>
</tr>
<tr>
<td>Graesser et al. (2007) study 2</td>
<td>50 minutes/ Once</td>
<td>Prior training and SEEK web tutor</td>
<td>Navigation only</td>
<td>Google™ search engine and websites</td>
<td></td>
</tr>
<tr>
<td>Hathorn &amp; Rawson (2012) study 1</td>
<td>30-40 minutes/ Once</td>
<td>1. Global monitoring 2. Adjunct inference questions</td>
<td>Text only</td>
<td>Text on computer, DirectRT</td>
<td></td>
</tr>
<tr>
<td>Hathorn &amp; Rawson (2012) study 2</td>
<td>30-40 minutes/ Once</td>
<td>1. Global monitoring 2. Specific monitoring</td>
<td>Adjunct fact questions</td>
<td>Text on computer, DirectRT</td>
<td></td>
</tr>
<tr>
<td>Johnson-Glenberg (2005)</td>
<td>2 weeks/ 4 times a week, 30 minutes per session</td>
<td>Metacognitive strategies: visual &amp; verbal</td>
<td>Anagram</td>
<td>3D-Reader</td>
<td></td>
</tr>
<tr>
<td>Kramarski &amp; Feldman (2000)</td>
<td>2 weeks (8 lessons)/NR</td>
<td>Metacognitive strategy in an internet classroom</td>
<td>Metacognitive strategy in a regular classroom</td>
<td>An internet environment</td>
<td></td>
</tr>
<tr>
<td>MacGregor (1988)</td>
<td>5-6 weeks/ 2-3 times per week 20 minutes per session</td>
<td>1. CTS + electronic dictionary 2. CTS + model literal questions 3. 1+2</td>
<td>Hard copies of reading materials</td>
<td>CTS</td>
<td></td>
</tr>
</tbody>
</table>
Note. NR = Not reported; ESL = English as a second language; EFL = English as a foreign language; ELL = English language learners; EO = English only; ERL: externally regulated learning; SRL: self-regulated learning; ICON: improving comprehension online; SDR: scaffolded digital reading; GALT: Glossing Authentic Language Texts; CTS: Computerized-mediated Text System; SAM-LS: social annotation model learning system; ULS: Universal Literacy Environment.

Design

The majority of studies (n = 14) employed a true experimental design (Cohen et al., 2011), randomly assigning students or classes to treatment or control conditions. Dreyer and Nel (2003) did not assign students randomly; therefore, it is considered to be a quasi-experimental study. Proctor, Dalton, and Grisham (2007) was a single-group design study without a control group, while Johnson-Glenberg’s (2005) was a within-subject design quasi-experimental study.

Participants

There were a total of 1,210 participants in the studies analyzed, excluding one study that did not report the number of participants (Mendenhall & Johnson, 2010). The majority of the participants (51%, n = 614) were undergraduate students. Participants from secondary schools numbered 412 (34%), and 184 (15%) were from elementary schools.

Participants’ native languages were usually unspecified (n = 10); however, from the descriptions of the studies, they were inferred to be native English speakers. Three studies stated that the participants were native English speakers (Hathorn & Rawson, 2012 Studies 1 & 2; Lomicka, 1998). Participants in Lomicka (1998) were learning French. Two studies included participants who spoke English only as well as bilingual speakers of English and other languages (Dalton et al., 2011; Proctor et al., 2007). Participants in Dreyer and Nel (2003) were speakers of African languages learning English. Participants in Kramarski and Feldman (2000) were learning English as a foreign language.

Duration and frequency

The instruction was implemented only once in more than half of the studies (n = 10). The duration varied across these studies, with the shortest instruction lasting between 30 and 40 minutes (Hathorn & Rawson, 2012 Studies 1 & 2) while the longest was 125 minutes (Gegner et al., 2009 Study 1).
The instruction in seven studies was implemented more than once (Dalton et al., 2011; Dreyer & Nel, 2003; Johnson-Glenberg, 2005; Kramarski & Feldman, 2000; MacGregor, 1988; Mendenhall & Johnson, 2010; Proctor et al., 2007). Notably, although implementing the instruction several times, Dreyer and Nel (2003) and Mendenhall and Johnson (2010) did not report frequency. The remaining five studies reported instruction frequencies ranging from 2-4 times per week. The average duration was 9.5 weeks (range = 2-13 weeks).

Findings

Findings by metacognitive strategy

The metacognitive strategies implemented by the 17 studies could be categorized into three main groups: (1) regulation, (2) strategy cues with think-aloud, and (3) vocabulary and comprehension support. We will address each group in the following.

There were another four studies focusing on the differences between students who read electronic texts and paper texts. Given this dissimilarity, we will report these studies separately (see Table 1).

Regulation as instruction

Metacognition has two fundamental aspects: knowledge of cognition and self-directed thinking (Johnson-Glenberg, 2005, p. 758). The instruction in seven studies (Azevedo et al., 2007; Graesser et al., 2007 Studies 1 & 2; Hathorn & Rawson, 2012 Studies 1 & 2; Johnson-Glenberg, 2005; Puntambekar & Stylianou, 2005 Study 2) involved elements related to the latter aspect.

Azevedo et al. (2007) compared the effects of externally regulated learning (ERL) and self-regulated learning (SRL) for undergraduate students learning science material via hypermedia. Participants in the ERL group received support from tutors, whereas participants in the SRL group did not. This support was for assisting students to achieve self-regulated learning. After reading the text, while both groups gained knowledge and understanding of the text, significant differences between ERL and SRL were found in labeling the human circulatory system and mental models of the system (ES = .55 and .48, respectively).

Similar to the support provided to the experimental group (i.e., the ERL group) in Azevedo et al. (2007), Hathorn and Rawson (2012, Study 1) intended to test if imbedded questions (i.e., global monitoring and adjunct inference) would promote self-monitoring and mental models in reading scientific text via a software program called DirectRT. The results indicated that asking global monitoring questions (i.e., “In the page you have just read, was anything different to what you thought? If so, why?”) maximized learners’ ability to put information on a diagram and reproduce concept maps of the content when compared to the text-only group without monitoring questions (ES = .79 and .73, respectively).

Puntambekar and Stylianou (2005, Study 2) gave participants metanavigation support (italics in original) to help them navigate and learn science content. As they defined it, metanavigation is “the support designed to enable students to reflect upon and monitor their link selection while navigating through a hypertext system” (p. 469). Hence, unlike Azevedo et al. (2007) and Hathorn and Rawson (2012, Studies 1 & 2), their metanavigation prompts included both questions (e.g., “What science concepts will help you solve the challenge?”) and suggestions (e.g., “Think about what other topic descriptions are related to the topic that you are reading”). Puntambekar and Stylianou (2005, Study 2) found that students who received metanavigation support outperformed students who did not in the concept-mapping test (ES = .61 for concept ratio and .70 for connection ratio).

Since the adjunct inference questions did not have an effect on students in the first study (ES = .21 for diagram and .05 for concept map), in their second study, Hathorn and Rawson (2012) sought to find out whether global monitoring questions or specific monitoring questions promoted memory of concepts and facts, and inference, when compared to the adjunct inference questions. Interestingly, they found that the monitoring questions did not have to be specific; participants in the global monitoring group still demonstrated a significant difference when they were asked inference questions compared to the specific monitoring group and the adjunct fact question group (ES = .94).
Unlike the previous four studies, instead of asking students questions, Johnson-Glenberg (2005) prompted students to ask questions and build mental models to monitor their understanding of the text and search for answers to their own questions. Utilizing a within-subject design, students first read texts with anagrams and then read texts with embedded verbal and visual support. The results indicated that students’ comprehension improved significantly ($ES = .45$) when reading texts with embedded support to generate questions and build mental models. Their re-reading behavior was also significantly increased with a moderate effect ($ES = .42$).

The two studies in Graesser et al. (2007) showed very different results. Basically, these two studies were designed to investigate if a web tutor can assist participants to discern the reliability of websites and learn content from websites. The experimental group in Study 1 had access to the SEEK web tutor (Source, Evidence, Explanation, and Knowledge), while the control group did not. The SEEK web tutor was designed to prompt students to think about and justify the reliability and trustworthiness of the websites provided on a mock Google™ page. The design was intended to encourage metacognitive skills of learning and self-regulated study. In the second study, Graesser et al. (2007) added prior training on critical stance to an experimental group to see whether providing such training in addition to the SEEK Tutor would strengthen the use of a critical stance among learners.

Surprisingly, the results of both studies were not very promising. In Study 1, even though the SEEK tutor group gained more knowledge and understanding than the control group, the effect was not significant ($ES = .37$, ns). Instead, participants in the control group outperformed the experimental group ($ES = 1.20$) on the process of studying the websites. The other measures were not affected by the imbedded support. The results of other measures in Study 2 remained the same as in Study 1.

Interestingly, however, the critical stance (where participants in the experimental group mentioned more critical thinking and causal relationships of why a volcano erupted) showed significant difference in favor of the SEEK web tutor group ($ES = .73$ for study 1 and $.57$ for study 2). Although the connection was not directly examined, Graesser et al. (2007) suggested that expressing critical stance information might have contributed to this difference.

**Strategy cues with think-aloud as instruction**

Annotations, one kind of strategy cue, appeared to provoke students to reflect more critically upon the primary text (Wolfe, 2008). Lomicka’s (1998) and Yanguas’s (2009) were the only two studies providing participants with strategy cues (i.e., glosses/annotations) with think-aloud as instruction. These two studies had commonalities. First, they both utilized strategy cues providing annotation via software. Additionally, they were designed to examine the differences among different types of glosses. Moreover, the participants in these two studies were English speakers who were learning French and Spanish, respectively.

However, the two studies produced very different results. The statistical analysis in Lomicka (1998) showed no significant difference between the three groups (all glosses, traditional glosses, no glosses) on the percentage of explanation stated by the 12 participants with think-aloud. Yanguas’s study (2009), which compared textual glosses, pictorial glosses, mixed glosses (experimental groups) and no glosses (control group), revealed significant differences in vocabulary recognition and comprehension between the experimental groups and control group. Large effect sizes (range $ES = 1.69$-$2.78$) indicated that multimedia annotation strongly supported vocabulary recognition and comprehension. Particularly, the mixed gloss group with textual and pictorial glosses outperformed all other groups in comprehension. Lomicka (1998) only collected think-aloud data from participants. If the study had utilized more rigorous vocabulary and comprehension measures, the result would possibly have looked more similar to that of Yanguas (2009).

In addition to the non-significant outcome, the software tracker data in Lomicka (1998) indicated that students in the full gloss group strongly preferred the traditional glosses (definition and translation), even though they could access other kinds of glosses (i.e., definition in French, images, references, questions, pronunciation, and translation in English). Lomicka (1998) argued that the multimedia annotations may assist reading comprehension and meta-comments, but the second language (L2) learners tended to only construct text-based comprehension, which led to a preference for traditional glosses.
Vocabulary and comprehension support as instruction

The third group of studies implemented instruction with both vocabulary and comprehension support (Dalton et al., 2011; Gegner et al., 2009 Studies 1 & 2; Proctor et al., 2007). The vocabulary and comprehension support in these four studies were either developing students’ metacognitive awareness or involving students in metacognitive processes.

Dalton et al. (2011) investigated which of the three strategies is the most effective given the Improving Comprehension Online (ICON) scaffolded digital reading (italics in original, SDR) environment: comprehension strategy only, vocabulary strategy only, and mixed strategies of comprehension and vocabulary. In a standardized measure using the Gates-MacGinitie comprehension subtest, there was no significant difference among these groups (ES = .08, ns).

Notably, when using researcher-developed measures, the mixed strategy group only outperformed the comprehension strategy group in the narrative comprehension measure (ES = .65). Yet, there was still no significant growth in expository comprehension among these three groups.

Still, all three groups showed significant growth in the Gates-MacGinitie vocabulary subtest (ES = .33). Additionally, the vocabulary only group and the mixed group both outperformed the comprehension only group in the researcher-designed vocabulary measure (ES = .58 and .96, respectively).

In Proctor et al. (2007), 30 fourth grade struggling readers (English only and English language learners) received a 4-week period of instruction embedded with both vocabulary and comprehension strategy support in a digital Universal Literacy Environment (ULE). Similar to Dalton et al. (2011), the result indicated no significant growth on the Gates-MacGinitie Comprehension pre- and post-tests (ES = .07). Yet, unlike Dalton et al. (2011), the participants showed no significant growth on the Gates-MacGinitie vocabulary pre- and post-tests either (ES = .18). The disaggregated data by English language learners and English only speakers showed similar trends with no significant gains from pre-test to post-test.

Nevertheless, considering that Gates-MacGinitie was a standardized test that might not be sensitive to the subtle change during a 4-week period of instruction, Proctor et al. (2007) chose to use features (i.e., number of clicks) provided by the digital ULE to calculate the correlation. By doing so, although not significant, they found that vocabulary gains were associated with number of clicks on the glossary.

The two studies in Gegner et al. (2009) investigated whether computer-based assistance on vocabulary and comprehension strategies (i.e., narrated animations, glossary terms, and motivational content) improved high school students’ comprehension, perceived difficulty, affect, and motivation of reading scientific articles. Differing from the results of the previous two studies, Gegner et al. (2009) found large effects on reading comprehension (ES = .79 for Study 1, ES = .82 for Study 2), and the outcomes were significantly different from those of the control groups (p < .0001). Furthermore, participants in the experimental groups considered that scientific articles are less difficult (ES = .96 for study 1, ES = .69 for study 2). These promising results might be due to multi-dimensional supports, such as visual, audio, vocabulary, and background knowledge.

Computerized environment versus hard copy

Lastly, four studies (24%) investigated the difference between reading on paper and reading in a computerized environment (Dreyer & Nel, 2003; Kramarski & Feldman, 2000; MacGregor, 1988; Mendenhall & Johnson, 2010). We consider that these studies investigated different aspects from previous studies as both experimental and control conditions were on the same computer-based interface. Hence, we grouped these studies together and discuss them in an individual section.

Kramarski and Feldman (2000) and MacGregor (1988) demonstrated similar results: the experimental groups did not outperform the control groups in the comprehension measures.
The experimental and the control groups in Kramarski and Feldman (2000) both received the same metacognitive strategy training (i.e., identifying the task, planning, performing and evaluation). In order to examine the contribution of an Internet environment, the experimental group was trained in the environment and the control group was trained in a regular classroom. The results were interesting; the participants in the experimental group did not outperform the control group in reading comprehension, reading strategies, or metacognitive awareness ($ES = -.27, -.48, -1.10$; respectively). They concluded that the Internet environment did not contribute significantly in terms of improving the students’ English reading comprehension, reading strategies, or metacognitive strategies. By giving the control group hard copy reading materials to examine the effects of a computerized-text system, similarly, MacGregor (1988) found no significant difference between the experimental groups and the control group ($ES = .50$ and $.07$, respectively) in any of the comprehension measures.

Interestingly, the experimental group in Kramarski and Feldman (2000) showed higher motivation to read than those in the control group ($ES = .73$). This phenomenon might be due to the novelty of the Internet environment. Likewise, MacGregor (1988) found that the experimental groups made more gains in two vocabulary tests than the control group when they had access to a computerized-mediated text system (CTS) with embedded electronic dictionary, model literal questions, or both ($ES = .88$). It is logical to believe that the experimental groups improved their vocabulary because they could look words up in the electronic dictionary, and benefited from the literal questions.

Mendenhall and Johnson (2010, Study 3) created a social annotation model learning system (SAM-LS) utilizing HyLighter to support students in cultivating critical thinking, writing, and related literacy skills for freshmen. The control group received hard copy reading materials instead of the HyLighter support. Again, the results indicated that there were no significant differences in measures of reading comprehension, critical thinking, or meta-cognition skills.

Since no significant differences between the two groups were found in any of the measures, to explain the unfavorable results, Mendenhall and Johnson (2010) further discussed the weakness of HyLighter that may contribute to the ineffectiveness of the outcome, such as the unstable environment, not fully capturing the intuition of human interaction, and users having to spend some time exploring its functions.

Dreyer and Nel (2003) showed different results from Mendehall and Johnson (2010, Study 3) and the previous two studies. The purpose of Dreyer and Nel’s study (2003) was to offer an academic course within a technology-enhanced environment. They used Varsite to support the instructional delivery in a university in South Africa. Varsite is a learning content management system with which instructors can create, store, manage, and deliver digital learning contents to learners. Learners can access an electronic study guide (with reading strategies) and other resources, perform assessments, and interact with peers via Varsite. In short, the participants in the experimental group received “strategic reading instruction” from this technology-enhanced environment.

The experimental group in Dreyer and Nel (2003) demonstrated a significant increase in English reading, reading pertaining to their profession (i.e., Communication), and TOEFL ($ES = .89, .78, .80$; respectively) compared with participants who did not have access to Varsite. They found that students who received the strategy instruction via the interactive learning system over 13 weeks greatly benefitted in terms of their reading comprehension and standardized tests.

**Findings by participant characteristics**

Studies included in this meta-analysis involved a wide range of participants. To answer the second research question, findings related to participant characteristics across studies of metacognitive instruction utilizing computerized environments are presented to offer a comprehensive understanding of the effects on participants of different grade level, reading ability, and language status.

**Grade level**

Students ranged in grade level from elementary school through university. To clearly present the effects of metacognitive strategy instruction in terms of participants’ grade level, first, we use the first three groups of studies
in the earlier section to discuss this issue. Then, we move on to those studies where computerized and non-computerized reading contexts were both employed.

First, in seven studies in which regulation was used as instruction, the participants were either undergraduate students (Azevedo et al., 2007; Graesser et al., 2007 Studies 1 & 2; Hathorn & Rawson, 2012 Studies 1 & 2) or 6th-7th graders (Johnson-Glenberg, 2005; Puntambekar & Stylianou, 2005, Study 2). Generally, the results were also very similar, revealing that regulation was an effective form of instruction for improving both older and younger students’ reading comprehension. The results of the two studies in Graesser et al. (2007) were an exception. We will address this phenomenon in more detail in the discussion.

Second, Dalton et al. (2011), Gegner et al. (2009, Studies 1 & 2) and Proctor et al. (2007) used both vocabulary and comprehension support. However, Gegner et al. (2009, Studies 1 & 2) aimed at high school students while the other two studies recruited 5th and 4th graders respectively. Generally speaking, the experimental groups in Gegner et al. (2009, Studies 1 & 2) greatly benefitted from the instruction (ES = .79 and .82 for two comprehension measures respectively). In contrast, the elementary students in Dalton et al. (2011) and Proctor et al. (2007) seemed not to improve their reading comprehension based on the instruction provided.

Then, Lomicka (1998) and Yanguas (2009) both used strategy cues with think-aloud as instruction to improve undergraduate students’ reading comprehension. Yet, as presented earlier, the two studies demonstrated very different results. Since the measure in Lomicka (1998) was not that rigorous, it might be problematic to state that the instruction used did not work on older students.

Lastly, students in the four studies which assigned students to either computerized or non-computerized reading contexts involved mixed age groups: elementary (3rd grade), middle school (8th grade) and undergraduate. Since the participants varied in their grade levels, it was challenging to determine the effects of these strategies according to their age. Essentially, only freshmen in Dreyer and Nel’s study (2003) showed significant growth.

Language status

Among seven studies which specifically reported participants’ language status, four (Dalton et al., 2011; Dreyer & Nel, 2003; Kramarski & Feldman, 2000; Proctor et al., 2007) included students who were ESL, EFL, ELL and/or bilinguals. We chose to focus on these studies to highlight the effects of metacognitive strategy instruction on students’ reading comprehension in terms of their language status.

Dalton et al. (2011) and Proctor et al. (2007) included students who spoke only English and those who were English-Spanish bilingual or other bilingual students. Proctor et al. (2007) analyzed students who spoke only English, and English language learners who spoke Spanish, and found that their pre-post comprehension growth after 4 weeks of instruction was not significantly different from each other (ES = .07). Dalton et al. (2011) found a similar trend, that is, there were no significant differences in the Gates-MacGinitie pre-post measures of comprehension between English only speakers and bilingual students who spoke languages other than English.

Notably, in a researcher-created measure on expository comprehension, Dalton et al. (2011) found that English only students outperformed bilingual students who spoke Spanish across conditions (ES = .36, p = .03). The effect might be influenced by English only students’ outstanding performance of vocabulary prior to the instruction.

In Dreyer and Nel (2003) and Kramarski and Feldman (2000), students were either ESL or EFL. As presented earlier, the control groups in both studies did not read in a computer-based environment. However, the findings differed greatly. Dreyer and Nel (2003) concluded that Varsite, a learning content management system, supported students’ learning, and their comprehension measures in English and other areas increased (ES = .78 – .89). At the opposite end, Kramarski and Feldman (2000) concluded that Internet access only helped increase students’ motivation, but not their comprehension, strategy use, or metacognitive awareness (ES = -.27 – -1.10).
Reading abilities

Six studies (35%) specified participants’ reading abilities. Gegner et al. (2009) recruited participants in AP (advanced placement) Biology classes. In contrast, Johnson-Glenberg (2005) recruited poor comprehenders, reading below the mean on the state’s standardized reading test or below the mean for the class in text comprehension. Participants in Proctor et al. (2007) were also struggling readers performing at the 23rd percentile in reading vocabulary and the 31st percentile in reading comprehension on the Gates–MacGinitie Reading Achievement Test.

MacGregor’s (1988) and Dreyer and Nel’s (2003) studies involved students with different reading abilities. Participants in the former were 3rd grade students reading at the 4th to 6th Stanine (average readers) or at the 6th Stanine (good readers). Dreyer and Nel (2003) disaggregated outcomes for successful readers (scoring above 55%) and at-risk readers (scoring below 55%).

A significant effect on the comprehension measure after instruction was observed in Gegner et al. (2009, Studies 1 & 2), indicating that comprehension aids in the form of narrated animations and glossary were effective for students in AP Biology classes (ES = .79 and .82, respectively). Dreyer and Nel (2003) had similar findings for their successful readers after using Varsite as a computer-based tool to deliver studying strategies and content (ES = 1.01–1.14). Yet, in MacGregor (1988), improvement of third grade students with good and average reading abilities was only observed in vocabulary measures, not in comprehension measures.

Essentially, the effects for at-risk or struggling students were less optimal. First of all, although struggling students in Proctor et al. (2007) improved in vocabulary and comprehension, none of them was significant (ES = .06 and .15). While participants in Johnson-Glenberg (2005) benefitted from the intervention of 3D-Reader, there was only a moderate effect (ES = .45). According to their further analysis of the at-risk readers, however, Dreyer and Nel (2003) found that those in the experimental groups outperformed the at-risk readers in the control groups in the English reading and TOEFL posttests (ES = .82 and .72), but not in content reading (Communication Reading) (ES = .30). As a side note, these effect sizes were directly cited from Dreyer and Nel (2003).

Findings by genre

Our third research question aimed at examining the effects of the metacognitive strategy instruction in terms of genre of instructional content because, as Lipson and Cooper (2002) pointed out, the skills needed to comprehend text vary by such factors as text form and genre. Therefore, in the following, we present our findings by different genres.

Except for Kramarski and Feldman (2000) and Mendenhall and Johnson (2010), the remaining studies all provided information about genre/text type. As a side note, the text was all written in English, except that Lomicka (1998) used a French poem and Yanguas (2009) used Spanish text.

Lomicka’s (1998) was the only study using literary text. The results indicated no significant difference between participants who read the poem in the experimental group (full glosses) and the control groups (traditional glosses, no glosses).

Expository texts were adopted by eleven studies. More specifically, all but one (i.e., MacGregor, 1988) employed scientific texts for reading. Most of them related to human biology (e.g., the heart and circulatory system) or natural science (e.g., volcanic eruptions). The effect sizes, although ranging from .07 to 2.72, were mostly moderate to large. Overall, the metacognitive strategy instruction appeared to be effective in helping students comprehend scientific texts. The instruction seemed to also benefit students in other aspects when reading scientific texts. For example, participants in Gegner et al. (2009, Studies 1 & 2) considered that scientific articles were less difficult. On the other hand, MacGregor (1988), who was the only one using non-scientific texts, found that the comprehension outcome did not reach a significant difference (ES = .50). A reasonable explanation was that the number of her participants was very small (n = 12).

In those studies with mixed expository/informational and narrative texts, the effects were also mixed. Dalton et al. (2011) found a moderate effect on narrative text comprehension (ES = .65), but a small and non-significant effect on expository text comprehension. Proctor et al. (2007) found a small and non-significant effect on narrative and
informational texts (ES = .07). Yet, Dreyer and Nel (2003), who specified that the texts were academic, showed promising results with communication and ESL reading comprehension (ES = .78-.89).

Findings by instruction and computerized environment analysis

To answer the final research question, we took six studies that had the most and the least effect sizes for the reading comprehension measures, and analyzed the instruction and type of computerized environment between the metacognitive strategy instructions and the control conditions.

Studies in which large effect sizes were observed incorporated glossary/vocabulary hyperlinks with definitions displayed by text or picture. Dreyer and Nel (2003) and Gegner et al. (2009, Studies 1 & 2) both implemented reading strategies and self-monitoring/self-check questions to facilitate students’ reading comprehension, while Yanguas (2009) focused solely on providing definitions in English or pictures of Spanish vocabulary. In general, despite the varying difficulties of the computerized interface, all three studies had large effects on reading comprehension.

The computerized interface of the three studies in which non-significant small effect sizes were observed were more complicated than the three studies described above. For example, Graesser et al. (2007, Study 1) asked participants to rate the reliability and relevance of the websites, and Mendenhall and Johnson (2010, Study 3) requested participants to highlight, take notes, and compare notes with peers via the HyLighter interface. Although these activities seemed to assist metacognitive thinking and monitoring, they did not target reading comprehension directly. On the other hand, Kramaski and Feldman (2000) targeted metacognition strategies and contextual clues that improved reading comprehension, but the control group received the same instruction. The only difference between the experimental and the control groups was that the experimental group had Internet access. Hence, the results only indicated that the use of a computerized interface is not a guarantee of improved comprehension.

Conclusions

Helping students to comprehend a variety of texts has always been a key issue for teachers (e.g., Vaughn et al., 2013). Recently, a growing body of research has started to focus on investigating reading comprehension processes while reading computer-based texts (e.g., Ertem, 2010). Many of these studies have especially aimed at examining the effects of the strategies provided to students who read digital texts (e.g., Henry, 2006).

The present meta-analysis chose to specifically examine the effects of metacognitive strategy instruction on students’ reading comprehension in computerized reading contexts. When reviewing the results of the 17 quantitative studies, however, the effects of these different examples of metacognitive strategy instruction remain mixed. Generally, some interesting trends may be drawn from this meta-analysis: 1) metacognitive strategy instruction seems to be more effective in assisting students in comprehending scientific texts, 2) regulation is deemed as a more effective form of instruction, 3) students who received instruction did not necessarily outperform their counterparts, and 4) students’ reading abilities played an interesting role.

Science texts are usually considered as more difficult than non-scientific texts. Hence, educators and researchers have put effort into investigating how to assist students in comprehending science readings such as the content of textbooks (e.g., Smith et al., 2010). Our findings are consistent with several previous studies (e.g., Spence, Yore, & Williams, 1999). In studies where science texts were used (e.g., Gegner et al., 2009, Study 2), larger effect sizes for measures of comprehension were found. Additionally, the metacognitive strategy instruction appeared to benefit students in aspects other than comprehension (e.g., motivation). This might not only reaffirm the impact of metacognitive strategies on helping students comprehend science readings, but also confirm that these strategies remain effective even when the text is presented in digital form.

As we stated in the introduction, showing teachers which form of metacognitive strategy instruction was more effective composed the core purpose of this study. According to our findings, regulation seemed to be more effective than other forms of instruction. Be it asking students questions to prompt them to monitor their reading (e.g., Azevedo et al., 2007) or prompting students to ask questions to monitor their reading (i.e., Johnson-Glenberg, 2005),
students were engaged in self-questioning. More specifically, regulation helped students to be more consciously aware of what they were reading; students who received this form of instruction therefore had better performance in the comprehension measures. Besides, although we found a more consistent result (i.e., regulation appeared to be an effective form of instruction for both younger and older students), little is known about whether the effects of metacognitive strategies are influenced by participants’ age. Therefore, future researchers might consider comparing students across different age groups as they receive metacognitive strategies to draw a more rigorous conclusion.

We also noticed an interesting phenomenon: no discernible differences in reading comprehension measures between the experimental groups were found in several studies (e.g., Dalton et al., 2011). In Graesser et al. (2007, Study 1), the control group even outperformed the experimental group in the process of studying the websites. One possible explanation for this occurrence is that these experimental groups’ cognitive load (Sweller, 1988) might have exceeded the capacity of their working memory. For example, in Graesser et al. (2007), the SEEK tutor was a relatively novel tool for the participants; meanwhile, the SEEK group was asked to conduct multiple actions including on-line ratings and structured note-taking tasks. Thus, Graesser et al. (2007) concluded, “Enhancements are needed in training quality, quantity, and/or both” (p. 103).

Yet, this might highlight a need to re-think the implementation as well as the design of computerized reading contexts. Although studies such as Kim (2013) found that the digital peer contributed to students’ performance in the immediate and delayed posttest text comprehension, there were studies which found totally different results. In the four studies which compared the differences between computerized and non-computerized reading contexts, three of them found that those who received the metacognitive strategy instruction from the computerized reading contexts did not outperform those who did not. In essence, it seems that the computerized reading contexts might not be the determining factor. For example, the HyLighter system in Mendenhall and Johnson (2010) emphasized social interactions among students. In other words, if the quality of students’ social interactions was not good enough, students’ reading comprehension might be affected. Combined, instructors’ roles or other factors such as peer interactions might outweigh the computerized reading contexts per se, just as Clark (2000) reminded us, “educational media do not cause learning, but rather educational methods cause learning” (cited in Gegner et al., 2009, p. 80).

The other intriguing finding was the effect of the instruction in terms of students’ reading abilities. As many studies indicated (e.g., Ertem, 2010), teaching poor or struggling readers strategies would be especially beneficial for them. Hence, it is assumed that less skilled readers would benefit more from instruction than those who already have sophisticated reading skills and/or strategies.

Nevertheless, according to our findings, struggling readers demonstrated mixed yet less optimal results. On the other hand, studies that recruited good or successful readers (e.g., Gegner et al., 2009) showed statistically optimal results as well as having much higher effect sizes. These results, which appear to be opposite to those of previous studies, are not that surprising. As Swanson and De La Paz (1998) and others (e.g., Hopkins & Mackay, 1997) have noted, proficient readers usually execute some metacognitive strategies or already employ other reading strategies. Thus, it might be easier for these skilled readers to quickly adopt or apply these new metacognitive strategies based on their existing reading strategies when they are instructed. In contrast, the task difficulty, the new text structure, and/or the relatively short duration of the instruction (see Proctor et al., 2007) might result in a less optimal or even non-significant result for poor/struggling readers.

The only study that reported at-risk students’ significant improvement (i.e., Dreyer & Nel, 2003) employed both computerized and non-computerized reading contexts. Hence, it is reasonable to assume that the strategies embedded in the computerized reading contexts might have contributed to the significant growth. In other words, with a well-designed computerized reading context, the effects of metacognitive strategy instruction on students’ reading comprehension might be enhanced. Nevertheless, as Ertem (2010) directly claimed, “we know very little about specifically which features of electronic text work best for struggling readers” (p. 142). More studies need to be done to ensure the complex interplay between metacognitive strategy instruction and computerized reading contexts, and how that improves poor/struggling readers’ reading comprehension.
Study limitations and implications

Our main limitation was the quantity of the studies included in this meta-analysis. This added difficulties to generalizing a united conclusion from 17 studies, especially because they varied in terms of the participants’ age, language status, foci of their strategies, etc.

Still, since this meta-analysis demonstrated mixed results in terms of the effects of various metacognitive strategies on comprehending non-traditional texts, this points out a need to conduct more empirical studies to determine what kinds of metacognitive strategies and what types of computerized reading contexts would have greater effects on students’ reading comprehension.

Besides, many of the included studies did not provide sufficient description of the participants. Yet, it appears that the role of students’ backgrounds deserves further attention when probing the effects of metacognitive strategy instruction on their reading comprehension. Thus, we urge researchers to provide a detailed description of participants, especially regarding their reading ability and language status, in future studies. By doing so, we believe that those who are interested in conducting a meta-analysis on related issues could draw a more solid conclusion from sufficient data.

The other implication for future research was inspired by our examination of the fourth group of instruction (i.e., Computerized Environment versus Hard Copy). As aforementioned, those metacognitive strategies did not enable participating students to have better comprehension of digital texts. Still, this tentative finding was made on the basis of very limited studies. More evidence on the effects of the same metacognitive strategy when students use electronic versus paper-based interfaces is needed.

The most important implication for practice was that teachers should know that no one instructional model could be recommended for all. Students’ various backgrounds, the characteristics of diverse content areas, and/or the designs of computerized reading contexts might have an influence on the effects of any specific reading comprehension strategy. As such, teachers at different levels should be more aware of the role of these variables and carefully use our findings as a resource in order to choose the “most effective” metacognitive strategy.

References


Appendix


What Drives Nurses’ Blended e-Learning Continuance Intention?

Yung-Ming Cheng

Department of Business Administration, Chaoyang University of Technology, No. 168, Jifeng E. Rd., Wufeng District, Taichung City 41349, Taiwan // ymcheng@mail.cyut.edu.tw

(Submitted January 20, 2013; Revised December 14, 2013; Accepted February 22, 2014)

ABSTRACT

This study’s purpose was to synthesize the user network (including subjective norm and network externality), task-technology fit (TTF), and expectation–confirmation model (ECM) to explain nurses’ intention to continue using the blended electronic learning (e-learning) system within medical institutions. A total of 450 questionnaires were distributed to nurses, of which, 352 (78.2%) questionnaires were returned, and 322 effective questionnaires were analyzed in this study, with an effective response rate of 71.6%. Collected data were analyzed using structural equation modeling. This study’s results strongly supported the extended ECM with all hypothesized links being significant. The results reveal that TTF makes the greatest impact on nurses’ blended e-learning continuance intention; hence, the blended e-learning system should be developed to fit with nurses’ work goals and needs to enhance their continued system usage intention by increasing the extent of their confirmation to the system. Besides, user network profoundly affects nurses’ intention to continue using the blended e-learning system, to boost nurses’ continued system usage intention, medical institutions should strive to encourage opinion leaders among nurses to have some influence on nurses’ belief and make them believe the system must be useful in its purpose.

Keywords

Nurses’ blended e-learning continuance intention, Expectation–confirmation model, User network, Task-technology fit, Structural equation modeling

Introduction

Currently, blended learning combining face-to-face and online elements provides learners with a hybrid learning environment (Delialioğlu, 2012), and it has been widely used in nursing education because of the flexibility afforded by this type of learning (Jonas & Burns, 2010; Smyth, Houghton, Cooney, & Casey, 2012). However, nurse staffing shortages and increased workload have gradually interrupted nurses’ family life and learning (Jonas & Burns, 2010; Tsang, Chen, Wang, & Tai, 2012), thus nurses still experience difficulty accessing further education even if they complete the courses via blended learning which combines a few days in authentic classrooms/laboratories and most time online. As compared to blended learning, blended electronic learning (e-learning) combines asynchronous and synchronous e-learning to provide learners with access to asynchronous and synchronous communication (Donnelly, 2010), thus it can assist instructors in achieving more effective teaching strategies and help for attracting more learners by offering better flexibility of the learning process (Zuvic-Butorac, Nebic, Nemcanin, Mikac, & Lucin, 2011). Hence, introducing blended e-learning in nursing education may further become a more flexible solution for nurses’ continuing education programs, especially for nurses with high workload and family commitments.

Essentially, there is increasing attention to blended e-learning as a flexible way of developing continuing education programs for nurses, and it has been accepted by its intended nurses (Hoffman et al., 2011; Jonas & Burns, 2010; Smyth et al., 2012). Noteworthily, while users’ initial acceptance of the information system (IS)/information technology (IT) is the first step towards its success, the eventual success of the IS/IT depends on its continued usage (Wu, 2013). However, far less emphasis has been placed on understanding whether nurses intend to continue using blended e-learning after having initially accepted it. Besides, prior studies (e.g., Liang, Wu, & Tsai, 2011; Tung & Chang, 2008) have explained nurses’ e-learning acceptance based on their utilitarian perceptions of the IS/IT, but simply focusing on nurses’ extrinsic motivators of e-learning may not be enough. Virtually, users’ IS/IT adoption may not be only determined by their utilitarian perceptions of the IS/IT but also by a good fit between their tasks and the IS/IT (Zhou, Lu, & Wang, 2010) and the influences of other users (Lee, 2010). Presently, the expectation–confirmation model (ECM), proposed by Bhattacharjee (2001a), is one of the most widely applied models in a variety of domains on continued IS usage (Kang, Hong, & Lee, 2009), and it can be used as the base for this study’s research model. Accordingly, this study’s purpose was to integrate user network and task-technology fit (TTF) as antecedents to fill the gaps of ECM in explaining nurses’ intention to continue using the blended e-learning system within medical institutions.
Theoretical background, hypotheses, and research model

ECM

Based on the expectation–confirmation theory (ECT) in the field of consumer behavior, Bhattacherjee (2001a) developed the ECM of IS continuance. ECM makes changes to the ECT by transforming the difference between pre-consumption expectation and experienced performance into a pure post-acceptance model in the field of IS continued use, and it posits that users’ intention to continue IS/IT usage is dependent on perceived usefulness (PU; i.e., post-adoption expectation), the extent of users’ confirmation, and users’ satisfaction with the IS/IT (Bhattacherjee, 2001a, 2001b). PU is defined as “the degree to which a person believes that using a particular system would enhance his/her job performance” (Davis, 1989, p. 320). Confirmation refers to the degree of users’ perception of the congruence between expectation of IS/IT use and its actual performance (Bhattacherjee, 2001a). In the e-learning context, the follow-up studies (e.g., Larsen, Sørebø, & Sørebø, 2009; Lee, 2010; Lin & Wang, 2012) derived from the ECM are validated using various field surveys of samples. First, learners’ PU of the e-learning system can be influenced by their confirmation of expectations (Larsen et al., 2009; Lee, 2010; Lin & Wang, 2012), and their PU has a positive effect on their intention to continue using the system (Lee, 2010; Lin & Wang, 2012). Next, learners’ satisfaction with the e-learning system is determined by their confirmation of expectations and their PU of the system (Lee, 2010; Lin & Wang, 2012). Finally, learners’ satisfaction with the e-learning system leads to their continued system usage intention (Larsen et al., 2009; Lee, 2010; Lin & Wang, 2012). Hence, this study hypothesizes:

• H1: Confirmation will positively affect PU of the blended e-learning system.
• H2: Confirmation will positively affect satisfaction of the blended e-learning system.
• H3: PU will positively affect satisfaction of the blended e-learning system.
• H4: PU will positively affect intention to continue using the blended e-learning system.
• H5: Satisfaction will positively affect intention to continue using the blended e-learning system.

Incorporating user network into ECM

Essentially, the user’s decision to accept a special type of system/technology may be influenced by other users, and such effects from user network can be classified into social influence and network externality (Pontiggia & Virili, 2010). User network also plays a key role in learning environments as an important source for social support (Cadima, Ojeda, & Monguet, 2012). In the e-learning context, two types of user network factors as the antecedents to ECM are inferred below.

Social influence profoundly affects user behavior (Ajzen, 1991; Lee, 2006), thus users’ decision to adopt a special type of IS/IT may usually be influenced by the suggestions of group leaders/other group members who have already used the IS/IT (Pontiggia & Virili, 2010). Essentially, the construct of “subjective norm” can capture such type of social influence (Ajzen, 1991; Fishbein & Ajzen, 1975). Subjective norm refers to an individual’s perception of the opinions of salient social referents about whether he/she should or should not perform the behavior in question (Fishbein & Ajzen, 1975). When learners perceive that their important referents think they should use the e-learning system, thus, they may incorporate their referents’ beliefs into their own belief, and they will further think the e-learning system must be useful in its purpose (Lee, 2006; Lee, Hsieh, & Ma, 2011; Van Raaij & Schepers, 2008), and tend to continue using the system with the influences of their salient referents (Lee, 2010). Hence, this study hypothesizes:

• H6: Subjective norm will positively affect PU of the blended e-learning system.
• H7: Subjective norm will positively affect intention to continue using the blended e-learning system.

Network externality refers to an increase in the utility of a product for a user as the number of other users of that product increases (Van den Ende, Wijnberg, Vogels, & Kerstens, 2003). Hence, the development of IS/IT is along with users’ perceptions of increasing numbers of others supporting that IS/IT; this will continuously add value to the IS/IT and further lead to increases in the number of other users’ IS/IT usage (Lee, 2006; Nault & Dexter, 1994; Wang & Seidmann, 1995). If learners perceive that increasing numbers of other learners are using the e-learning system, they will think the system must be useful in its purpose and try out the system (Lee, 2006). Hence, this study hypothesizes:

• H8: Network externality will positively affect PU of the blended e-learning system.
• H9: Network externality will positively affect intention to continue using the blended e-learning system.
Combining TTF with ECM

TTF model extends the technology acceptance model (TAM) by considering how the task affects usage intention (Klopping & McKinney, 2004), and it postulates that technology acceptance depends on how well the new technology fits the needs of a particular task (Goodhue & Thompson, 1995). TTF refers to a matter of how the capabilities of the IS/IT match the tasks that the user must perform (Goodhue & Thompson, 1995; Larsen et al., 2009). Recently, TTF has been widely used and combined with other models such as ECM or TAM to explain the IS/IT adoption (Dishaw & Strong, 1999; Larsen et al., 2009; Zhou et al., 2010). Larsen et al. (2009) combined the ECM with the view of TTF to explore teachers’ acceptance of the e-learning tool in university colleges, and they found that the more the e-learning tool met teachers’ specific task characteristics, the higher was the probability that the e-learning tool would enhance teachers’ PU of the tool. Lin and Wang (2012) used the IS success model, TTF, and ECM to investigate learners’ continuance intentions within the blended learning instruction, and they showed that TTF both in terms of utilizing the e-learning system and supporting learning requirements resulted in the confirmation of system usage. Further, Lin (2012) developed a hybrid model by integrating IS continuance theory with TTF to explore the antecedents of the continuance intentions of virtual learning system (VLS) within the university, and further showed that TTF significantly and positively led to satisfaction and continuance intention of the VLS. Hence, this study hypothesizes:

- H10: TTF will positively affect PU of the blended e-learning system.
- H11: TTF will positively affect confirmation of the blended e-learning system.
- H12: TTF will positively affect satisfaction of the blended e-learning system.
- H13: TTF will positively affect intention to continue using the blended e-learning system.

Research model

This study’s research model incorporates the views of user network and TTF into the ECM. Based on the ECM, the research model presents three antecedents (i.e., subjective norm, network externality, and TTF) that lead to nurses’ intention to continue using the blended e-learning system. The research model used in this study is depicted in Figure 1.

![Figure 1. The research model](image-url)
Methods

Measurement and pre-test

In this study, responses to the items in subjective norm, network externality, TTF, confirmation, PU, satisfaction, and continuance intention were measured on a 7-point Likert scale from 1 (= “strongly disagree”) to 7 (= “strongly agree”) with 4 labeled as neutral. Items chose for the constructs in this study were adapted and revised from previous research. The questionnaire was pre-tested on 38 nurses from one hospital in Taiwan in the voluntary and anonymous way. This selected hospital had implemented the blended e-learning system at least one year ago, and the nurses from this selected hospital were using the blended e-learning system in their learning. Nurses were asked to identify any ambiguities in the meanings, and the questionnaire was revised based on their comments. The instrument’s reliability was evaluated, and the Cronbach’s \( \alpha \) values (ranging from .80 to .96) exceeded common requirements for exploratory research, indicating a satisfactory reliability level (Hair, Anderson, Tatham, & Black, 1998; Nunnally, 1978). The nurses who had participated in the pre-test were excluded from the final data collection and subsequent study. The final items are listed in Table 1 along with their sources.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
<th>Measure</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective norm (SN)</td>
<td>SN1</td>
<td>People who are important to me think that I should use the blended e-learning system.</td>
<td>Lee (2010)</td>
</tr>
<tr>
<td></td>
<td>SN2</td>
<td>People who influence my behavior think that I should use the blended e-learning system.</td>
<td>Van Raaij and Schepers (2008)</td>
</tr>
<tr>
<td></td>
<td>SN3</td>
<td>People whose opinions I value think that I should use the blended e-learning system.</td>
<td></td>
</tr>
<tr>
<td>Network externality (NE)</td>
<td>NE1</td>
<td>Most nurses in my department use the blended e-learning system.</td>
<td>Chun and Hahn (2007)</td>
</tr>
<tr>
<td></td>
<td>NE2</td>
<td>Most nurses in my hospital use the blended e-learning system.</td>
<td>Lee (2006)</td>
</tr>
<tr>
<td></td>
<td>NE3</td>
<td>As more and more nurses use the blended e-learning system, I think related services and supports will soon be developed.</td>
<td></td>
</tr>
<tr>
<td>Task-technology fit (TTF)</td>
<td>TTF1</td>
<td>Using the blended e-learning system fits with my work goals and needs.</td>
<td>Goodhue and Thompson (1995)</td>
</tr>
<tr>
<td></td>
<td>TTF2</td>
<td>Using the blended e-learning system fits with the way I like to enhance the efficiency of my work.</td>
<td>Larsen et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>TTF3</td>
<td>Using the blended e-learning system fits with the way I like to strengthen my professional skills in working within many clinical fields.</td>
<td>Lin (2012)</td>
</tr>
<tr>
<td></td>
<td>TTF4</td>
<td>Using the blended e-learning system is helpful for my work.</td>
<td></td>
</tr>
<tr>
<td>Confirmation (Conf)</td>
<td>Conf1</td>
<td>My experience with using the blended e-learning system was better than I expected.</td>
<td>Bhattacherjee (2001a)</td>
</tr>
<tr>
<td></td>
<td>Conf2</td>
<td>The service level provided by the blended e-learning system was better than I expected.</td>
<td>Bhattacherjee (2001b)</td>
</tr>
<tr>
<td></td>
<td>Conf3</td>
<td>My expectations from using the blended e-learning system were confirmed.</td>
<td>Larsen et al. (2009)</td>
</tr>
<tr>
<td>Perceived usefulness (PU)</td>
<td>PU1</td>
<td>Using the blended e-learning system enhances my learning effectiveness.</td>
<td>Lee (2006)</td>
</tr>
<tr>
<td></td>
<td>PU2</td>
<td>Using the blended e-learning system gives me greater control over learning.</td>
<td>Liang et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>PU3</td>
<td>I find the blended e-learning system to be useful in my learning.</td>
<td></td>
</tr>
<tr>
<td>Satisfaction (Satisf)</td>
<td>Satisf1</td>
<td>I am satisfied with the performance of the blended e-learning system.</td>
<td>Bhattacherjee (2001a)</td>
</tr>
<tr>
<td></td>
<td>Satisf2</td>
<td>I am pleased with the experience of using the blended e-learning system.</td>
<td>Bhattacherjee</td>
</tr>
</tbody>
</table>
learning system.

I am happy with the functions provided by the blended e-

Satisf3 (2001b)

Lee (2010)

Continuance intention

CI1 I intend to continue using the blended e-learning system

Bhattacherjee (2001a)

CI2 I will use the blended e-learning system on a regular basis

Bhattacherjee (2001b)

CI3 I will frequently use the blended e-learning system in the

Lin and Wang (2012)

future.

Sample

The learning management system (LMS) has been one of the most commonly used types of e-learning systems in institutions (Hershkovitz & Nachmias, 2011; McGill & Klobas, 2009). Hence, hospitals were selected by using two criteria: (1) the hospitals must have implemented the same blended e-learning system, the focus in this study is on the LMS with a blend of asynchronous and synchronous technologies; and (2) to ensure experience among nurses, hospitals must have used the LMS with a blend of asynchronous and synchronous technologies for more than one year prior to the study period, and nurses from these hospitals must have experience in using the LMS with a blend of asynchronous and synchronous technologies. This study’s sampling frame was taken from among nurses working in hospitals with over 500 beds in Taiwan. Initial phone calls were made to each hospital to confirm whether they matched this study’s selection criteria; if they matched the selection criteria, then this study would further explain the study’s purpose and invite their participation. Overall, three hospitals in Taiwan matching the selection criteria agreed to participate in this study.

Data collection and ethical considerations

Each hospital that participated in this study was asked to identify a contact person who could distribute the survey questionnaires to nurses who had experience in using the LMS with a blend of asynchronous and synchronous technologies. Before the distribution of the questionnaires, further phone calls were made to the designated contact person from each hospital to ask them to remind the nurses to read the study’s purpose and respondents’ rights that stated in the questionnaire cover letter. Each hospital was given 150 questionnaires and those were distributed to nurses who had experience in using the LMS with a blend of asynchronous and synchronous technologies. A total of 450 questionnaires were distributed to nurses in three hospitals, of which, 352 nurses agreed to participate in this study. Before answering the questionnaire, these nurses were asked to read the questionnaire cover letter that explained this study’s purpose, informed the voluntary and anonymous nature for their responses, and notified the right to withdraw from participation in this study. Return of the nurses’ completed questionnaire and their permission implied consent to participate in this study, and the obtained data from the nurses were used only for this study. Finally, 352 (78.2%) questionnaires were returned, 30 of these received questionnaires were discarded due to partial portions of missing values; consequently, 322 effective questionnaires were analyzed in this study, with an effective response rate of 71.6%.

Data analysis

This study’s data analysis followed a two-step method for structural equation modeling (SEM) approach recommended by Anderson and Gerbing (1988). First, confirmatory factor analysis (CFA) was used to develop the measurement model. Second, to explore the causal relationships among all constructs, the structural model for the research model was tested by using SEM. The statistical analysis software packages used to perform these analyses were AMOS 5.0 (SPSS, Inc., Chicago, IL, USA) and SPSS 8.0 (SPSS, Inc., Chicago, IL, USA).
Results

Descriptive characteristics of the effective respondents

A total of 322 effective responses were collected. All respondents were female, among them, 125 (38.8%) had less than 5 years of work experience, 102 (31.7%) had 5-10 years, 68 (21.1%) had 11-15 years, 17 (5.3%) had 16-20 years, and 10 (3.1%) had above 20 years. The majority of respondents (79.5%, \( n = 256 \)) had graduated from university or above. 211 (65.5%) were licensed practical nurses, and 111 (34.5%) were registered nurses. Besides, 37 (11.5%) were supervisors, and 285 (88.5%) were non-supervisors. Descriptive characteristics of the effective respondents are depicted in Table 2.

Table 2. Descriptive characteristics of the effective respondents (\( N = 322 \))

<table>
<thead>
<tr>
<th>Demographics</th>
<th>( n )</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Work experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>under 5 years</td>
<td>125</td>
<td>38.8</td>
</tr>
<tr>
<td>5-10 years</td>
<td>102</td>
<td>31.7</td>
</tr>
<tr>
<td>11-15 years</td>
<td>68</td>
<td>21.1</td>
</tr>
<tr>
<td>16-20 years</td>
<td>17</td>
<td>5.3</td>
</tr>
<tr>
<td>above 20 years</td>
<td>10</td>
<td>3.1</td>
</tr>
<tr>
<td><strong>Educational level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associate’s degree</td>
<td>66</td>
<td>20.5</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>238</td>
<td>73.9</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>18</td>
<td>5.6</td>
</tr>
<tr>
<td><strong>Type of nurse</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Licensed practical nurses</td>
<td>211</td>
<td>65.5</td>
</tr>
<tr>
<td>Registered nurses</td>
<td>111</td>
<td>34.5</td>
</tr>
<tr>
<td><strong>Position</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisors</td>
<td>37</td>
<td>11.5</td>
</tr>
<tr>
<td>Non-supervisors</td>
<td>285</td>
<td>88.5</td>
</tr>
</tbody>
</table>

*Note. All respondents in this study were female.*

Results of structural modeling analysis

Measurement model

To assess the measurement model, three analyses were conducted in this study. First, with regard to reliability, according to the views of previous studies (e.g., Hair et al., 1998; Holmes-Smith, 2001; Nunnally, 1978), squared multiple correlation (SMC) for each item, and composite reliability (CR) and average variance extracted (AVE) for each construct were used in this study to test the reliability of all constructs. The results of CFA showed that the SMC values for all items were greater than .50, which indicated a good reliability level (Holmes-Smith, 2001). The values of CR and AVE for all constructs exceeded the minimum acceptable values of .70 and .50 (Hair et al., 1998; Holmes-Smith, 2001; Nunnally, 1978), indicating a good reliability level and subsequently yielding very consistent results. Hence, the results of CFA demonstrated an acceptable level of reliability for all constructs. Moreover, the reliability coefficients of all constructs assessed by the Cronbach’s \( \alpha \) coefficient exceeded the .70 cut-off value as recommended by Hair et al. (1998) and Nunnally (1978). The results of reliability test are shown in Table 3.

Table 3. Results of CFA, validity analysis, and reliability test

<table>
<thead>
<tr>
<th>Construct item</th>
<th>Estimate</th>
<th>( t )-value</th>
<th>Standardized path coefficients</th>
<th>SMC</th>
<th>CR</th>
<th>AVE</th>
<th>Cronbach’s ( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN</td>
<td></td>
<td></td>
<td></td>
<td>.83</td>
<td>.62</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>SN1</td>
<td>1.00</td>
<td>— (^a)</td>
<td>.75</td>
<td>.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN2</td>
<td>1.14</td>
<td>13.26</td>
<td>.82</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN3</td>
<td>.95</td>
<td>13.21</td>
<td>.81</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NE</td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
<td>.79</td>
<td>.90</td>
<td></td>
</tr>
</tbody>
</table>
Second, in regard to validity, logical dimensions of validity (i.e., content validity and face validity) and statistical dimensions of validity (i.e., convergent validity and discriminant validity) were used in this study to assess the validity of all constructs. Content validity ensures that construct items are drawn from a review of relevant literature (Cronbach, 1951). Essentially, items chose for the constructs in this study were adapted and revised from previous research (see Table 1), where they had been shown to exhibit strong content validity. Face validity refers to whether respondents perceive the construct items to be applicable and credible (Cronbach, 1971). The questionnaire was pre-tested on 38 nurses from one hospital in Taiwan in the voluntary and anonymous way, and the questionnaire was revised based on their comments. Hence, items chose for the constructs in this study had strong face validity. As for convergent validity, based on Anderson and Gerbing’s (1988) rule, the results of CFA showed that the $t$-value of every item exceeded the 1.96 value ($p < .05$), so the evidence of good convergent validity was obtained as the items represented their constructs significantly (see Table 3). Finally, to test for discriminant validity, the procedure described by Fornell and Larcker (1981) was used in this study. The results of CFA showed that the AVE of each construct was distinct (see Table 3, Table 4).

Third, the most common rules used in performing the CFA for measurement model and testing the structural model include stipulating that the goodness-of-fit index (GFI) should be greater than .90, the adjusted GFI (AGFI) should be greater than .80, the incremental fit index (IFI) should be greater than .90, the Tucker-Lewis index (TLI) should be greater than .90, the comparative fit index (CFI) should be greater than .90, the root mean square error of approximation (RMSEA) should be less than .08, and the $\chi^2/df$ should be less than 3 (Adams, Nelson, & Todd, 1992; Bagozzi & Yi, 1988; Hair et al., 1998). The overall fit indices of measurement model were $\chi^2 = 269.18$, $df = 188$, $\chi^2/df = 1.43$, $p < .001$, GFI = .93, AGFI = .90, IFI = .99, TLI = .98, CFI = .99, and RMSEA = .04. Thus the results of CFA showed that the indices were over their respective common acceptance levels.

### Table 4. Discriminant validity for the measurement model

<table>
<thead>
<tr>
<th>Construct</th>
<th>SN</th>
<th>NE</th>
<th>TTF</th>
<th>Conf</th>
<th>PU</th>
<th>Satisf</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN</td>
<td>.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NE</td>
<td>.05</td>
<td>.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $a$ The loading was fixed.
Note. The italic values along the diagonal line are the AVE values for the constructs, and the other values are the squared correlations for each pair of constructs.

Structural model

The further step is to test the structural model for the research model depicted in Figure 1. The overall fit indices for the structural model were as follows: \( \chi^2 = 361.64, df = 196, \chi^2/df = 1.85, p < .001, GFI = .91, AGFI = .88, IFI = .97, TLI = .96, CFI = .97, \) and RMSEA = .05. Based on the rules of prior studies (e.g., Adams et al., 1992; Bagozzi & Yi, 1988; Hair et al., 1998), the results of CFA showed that the fit indices for this structural model were quite acceptable.

Hypothesis testing

Properties of the causal paths, including standardized path coefficients \((\beta)\) and \(t\)-values, are shown in Figure 2. As for antecedents to nurses’ blended e-learning acceptance, subjective norm had significant effects on PU \((\beta = .28, p < .001)\) and continuance intention \((\beta = .24, p < .001)\); hence, H6 and H7 are supported. Network externality had significant effects on PU \((\beta = .31, p < .001)\) and continuance intention \((\beta = .13, p < .05)\); hence, H8 and H9 are supported. TTF had significant effects on PU \((\beta = .24, p < .001)\), confirmation \((\beta = .34, p < .001)\), satisfaction \((\beta = .12, p < .05)\), and continuance intention \((\beta = .16, p < .01)\); hence, H10, H11, H12, and H13 are supported. As to nurse beliefs and continuance intention, first, confirmation had significant effects on PU \((\beta = .23, p < .001)\) and satisfaction \((\beta = .41, p < .001)\), and PU had a significant effect on satisfaction \((\beta = .13, p < .05)\); hence, H1, H2 and H3 are supported. Next, the effects of PU \((\beta = .28, p < .001)\) and satisfaction \((\beta = .33, p < .001)\) on continuance intention were significant; hence, H4 and H5 are supported. Further, using the empirical results above, the direct and indirect effects between the constructs are shown in Table 5.

Note. 1. Standardized path coefficients \((\beta)\) are reported \((t\)-values in parentheses).  
2. Absolute \(t\)-value > 1.96, \(p < .05\); absolute \(t\)-value > 2.58, \(p < .01\); absolute \(t\)-value > 3.29, \(p < .001\).
Table 5. Direct and indirect effects between the constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Conf</th>
<th>PU</th>
<th>Satisf</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DE</td>
<td>InDE</td>
<td>TE</td>
<td>DE</td>
</tr>
<tr>
<td>SN</td>
<td>-</td>
<td>-</td>
<td></td>
<td>.28</td>
</tr>
<tr>
<td>NE</td>
<td>-</td>
<td>-</td>
<td></td>
<td>.31</td>
</tr>
<tr>
<td>TTF</td>
<td>.34</td>
<td>-</td>
<td>.34</td>
<td>.24</td>
</tr>
<tr>
<td>Conf</td>
<td>-</td>
<td>-</td>
<td></td>
<td>.23</td>
</tr>
<tr>
<td>PU</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Satisf</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

Note. DE = direct effects; InDE = indirect effects; TE = total effects.

Discussion and implications

As expected, this study’s results strongly supported the ECM with all hypothesized links being significant. This study confirmed that nurses’ levels of confirmation and PU were key determinants of their levels of satisfaction, confirmation also had a significant impact on PU, and confirmation had the dominant effect on satisfaction; further, both satisfaction and PU were significant determinants of nurses’ continuance intention of the blended e-learning system, and the saliency of satisfaction was much stronger than PU. The results reveal the saliency of the ECM in understanding nurses’ blended e-learning continuance intention. The results further implicate that having satisfied users is the key driver of continued IS usage intention because they place more emphasis on the confirmation of their expectations in forming their levels of satisfaction than on their post-adoption belief (Bhattacherjee, 2001a, 2001b). Thus, the inclusion of users’ satisfaction in understanding the e-learning system usage continuance is strongly warranted (Larsen et al., 2009; Lin & Wang, 2012). Practically, Smyth et al. (2012) showed that the blended e-learning offered rich virtual learning environments in which interactions occurred among nurses asynchronously and synchronously, nurses might get expected benefits via their usage experiences with the blended e-learning, and thus it could enhance nurses’ satisfaction and continuance intention of the blended e-learning. Hence, this study’s results fit in with the foregoing views.

In regard to social antecedents to nurses’ blended e-learning acceptance, first, subjective norm was found to significantly affect nurses’ PU of the blended e-learning system and their continuance intention of the system. The result reveals that nurses internalize the opinions of their important referents. Further, the result implicates that such internalization may take place gradually (Van Raaij & Schepers, 2008), and the opinions of users’ important referents may gradually become part of their belief structure (Van Raaij & Schepers, 2008; Venkatesh & Davis, 2000). Hence, to make nurses’ learning via the blended e-learning system useful and further enhance their intention to continue using the system, this study suggests that medical institutions should take notice of the effects of social influences. Concretely speaking, medical institutions may utilize the effects of social influences (e.g., interpersonal communication, expert opinions, news reports, and mass media reports, etc.) to create the positive feedback loop for nurses’ perception of usefulness of the blended e-learning system, and to further facilitate their intention to continue using the system. Accordingly, internalization may take place regardless of whether users’ acceptance of the e-learning system is mandatory or voluntary (Van Raaij & Schepers, 2008). Next, network externality was also found to significantly affect nurses’ PU of the blended e-learning system and their intention to continue using the system. The result reveals that learners’ perception of the e-learning system acceptance by other users can create bandwagon effects (Lee, 2006), that is, network externality can just be the typical driver of bandwagon effects among nurses, and such effects will be rapidly strengthened as more nurses participate in e-learning activities. The result implicates that medical institutions should be conscious of the benefits for the phenomenon of network externality and further utilize such a phenomenon to entice more nurses to try out the blended e-learning system. In order to boost nurses’ blended e-learning continuance intention, this study suggests that medical institutions should strive to encourage opinion leaders among nurses, who have already used the blended e-learning system, to have some influence on nurses’ belief, and make these nurses believe the system must be useful in its purpose. Prior studies (e.g., Jonas & Burns, 2010; Smyth et al., 2012) indicated that blended e-learning might create a space where nurses could engage in supportive discussion surrounding their learning and allow them to socially interact with their peers, and instructors as facilitators of blended e-learning might further make nurses feel confident and enable them to feel connected to...
each other; these could enhance nurses’ ability to construct new ideas and receive feedback via dialogues with their instructors and peers. Therefore, this study’s results can explain the foregoing views.

With regard to TTF and nurses’ blended e-learning acceptance, TTF was found to significantly affect nurses’ PU, confirmation, satisfaction, and continuance intention. First, the result reveals that TTF contributes significantly to PU, confirmation, and satisfaction, which together explain nurses’ intention to continue using the blended e-learning system; further, the result implicates that TTF plays an important role as an antecedent to confirmation but less important to PU and satisfaction, and it indirectly affects nurses’ intention to continue using the blended e-learning system mainly through confirmation and secondarily through PU and satisfaction. Next, the result reveals that satisfaction, PU, and TTF respectively exhibit significantly positive impacts on nurses’ intention to continue using the blended e-learning system. However, the result implicates that nurses intend to continue using the blended e-learning system mainly because they satisfy their system usage and secondarily because they perceive the system to be useful to their learning and think the system is a good fit for the tasks that it support. Synthetically speaking, TTF reveals stronger indirect impacts on nurses’ intention to continue using the blended e-learning system than its direct impact. Hence, this study suggests that the blended e-learning system should be developed to fit with nurses’ work goals and needs to stimulate their confirmation based on a set of their initial expectations to the system, thus the system can in turn deliver satisfaction and benefits of system usage to nurses and further enhance their continued usage intention by increasing the extent of their confirmation to the system. Actually, Hoffman et al. (2011) revealed that nurses were highly satisfied with the clinical scenarios when using blended e-learning, because they believed it was a useful way to fit with their work goals, and it could engage in authentic clinical learning and develop cognitive skills such as clinical reasoning, problem-solving and decision-making. Unquestionably, this study’s results just mirror the foregoing views.

Conclusions

This study proposes an extended ECM with the inclusion of the roles for user network and TTF to explain nurses’ blended e-learning continuance intention, and this extended ECM is strongly supported by empirical evidence, with all significant paths in the hypothesized directions. The following contributions are particularly worth mentioning. First, the application of the ECM with the view of user network in the field of nurses’ e-learning continuance intention reveals deep insights when facing the problem of bandwagon effects among nurses. That is, users may feel obligated to participate because they may want to belong to a user community, and such a community will facilitate their IS continuance intention (Hsu & Lu, 2004; Pontiggia & Virili, 2010). Nevertheless, previous studies in nurses’ acceptance of the e-learning system do not consider focusing more on exploring the effects of user network factors (i.e., subjective norm and network externality) not only on users’ initial usage, but also on the second phase of the acceptance process leading to their continued system usage, which are the key antecedents of post-adopter belief (i.e., PU). Hence, this study’s empirical evidence contributes significantly to the body of nursing research on bridging the gap of limited evaluation for the effects of user network on nurses’ e-learning continuance intention which is very scarce in an ECM aspect. Next, it should be noted that the work-centric factor as TTF is focused more on users’ perception for a good fit between their tasks and the IS/IT (Larsen et al., 2009; Zhou et al., 2010), thus this study contributes to an understanding of the TTF in explaining nurses’ blended e-learning continuance intention that is difficult to explain with only their utilitarian perception of the system.

Several limitations and suggestions for further research should be noted in this study. First, all respondents in this study are female because women still make up the overwhelming majority of nurses in Taiwan, and thus this study’s results tend to model the specific behavior of female nurses rather than general behavior of all nurses. Prior studies have shown that gender differences can cause discrepancies in the effects of subjective norm, PU, and attitude on users’ behavioral intention (Armitage, Norman, & Conner, 2002; Ong & Lai, 2006). Accordingly, further research may examine the moderating effect of gender on the relationship between nurses’ belief and their blended e-learning continuance intention. Second, all respondents in this study have experience in using the LMS with a blend of asynchronous and synchronous technologies. Further research should investigate and compare the continuance behavior of blended e-learning between nurses who are familiar with and who are not familiar with the blended e-learning to generalize this study’s findings. Finally, this study does not consider previous exposure to technology (PET) as a factor which may influence the IS/IT acceptance among users who are not so enthusiastic about using the
IS/IT (Holzinger, Searle, & Wernbacher, 2011). Taking into account PET construct may enrich the model of nurses’ continued blended e-learning usage intention.

References


The impacts of social network structure and work with positive externalities.


Exploring Students’ Knowledge Construction Strategies in Computer-Supported Collaborative Learning Discussions Using Sequential Analysis

Nurbiha A Shukor*, Zaidatun Tasir, Henny Van der Meijden and Jamalludin Harun
Faculty of Education, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia // Faculty of Education, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia // Radboud University, 6525 Nijmegen, Netherlands // Faculty of Education, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia // nurbiha@utm.my // p-zaida@utm.my // h.vandermeijden@pwo.ru.nl // p-jamal@utm.my

*Corresponding author

(Submitted April 11, 2013; Revised December 04, 2013; Accepted January 13, 2014)

ABSTRACT
Online collaborative learning allows discussion to occur at greater depth where knowledge can be constructed remotely. However students were found to construct knowledge at low-level where they discussed by sharing and comparing opinions; those are inadequate for new knowledge creation. As such, this research attempted to investigate the students’ behaviour and their strategies to construct knowledge during online collaborative discussions. Using the combination of content analysis and sequential analysis technique, this research found that groups those being able to construct high-level knowledge tend to negotiate on shared information. Argumentation is also found to contribute for successful knowledge construction at higher-level. This study suggests triggering argumentation and emphasizing on problem-solving tasks for better knowledge construction sustainability.

Keywords
Knowledge construction, Sequential analysis, CSCL, Online learning discussion, Higher education

Introduction
Knowledge construction develops in a collaborative learning environment where students communicate by sharing information in groups for solving given tasks (Dillenbourg & Fischer, 2007; Alavi & Dufner, 2005; Crook, 1998). Knowledge construction per se, is the vivid evidence of collaborative learning taking place (Alavi & Dufner, 2005; Veerman & Veldhuis-Diermanse, 2001), as students learning in a collaborative learning environment have to make their thoughts clear (Van Boxtel, 2000).

Previous studies found that students prefer to share and compare the available information rather than progressing to construct new knowledge during collaborative discussions (Ma, 2009; Schellens et al., 2008). It shows that students tend to interact at the level of rapid consensus, where students tend to accept peers’ opinions not necessarily because they agree with each other, but merely to hasten the discussion (Rimor, Rosen & Naser, 2010).

Constructing knowledge at a higher level is more important for students’ learning, particularly online, because it ensures students are experiencing meaningful learning. At the high-level of knowledge construction, students externalize thought that involve arguments, justification, or decision making. Those are the attributes that help students to be critical thinkers and thereby able to construct new knowledge (McLoughlin & Luca, 2000).

Online collaboration often involves the use of discussion board as the platform to work collaboratively. Using this platform, students can negotiate opinions with group members to finally construct knowledge (Schellens et al., 2008; Beers et al., 2005; Van der Meijden, 2005). Researches indicate that students communicating in asynchronous medium posted more messages of a higher-level of knowledge construction as compared to messages in synchronous communication (Van der Meijden, 2005; Veerman & Veldhuis-Diermanse, 2001). This is because asynchronous communication provides retention time for; self-reflection (such as the time to provide opinions and reflecting information) (Veerman & Veldhuis-Diermanse, 2001), processing information (Kim, Liu & Bonk, 2005) and being aware of how group dynamics evolve (Solimeno et al., 2008).

In this study, students’ behaviour of constructing knowledge will be investigated. Upon knowing their knowledge construction behaviours, a more detailed exploration will be carried out to know how they have progressed in constructing knowledge; that is, what are the strategies they used to construct knowledge. From this information, the
educators will be able to know which strategy is significant that would help the students to be able to construct knowledge particularly towards high level. As such, the following are the research objectives:

- To investigate students’ behaviour of constructing knowledge in online collaborative discussions
- To identify students’ strategies when constructing knowledge in online collaborative discussions
- To identify the transition state of students’ knowledge construction strategies in online collaborative discussions

**Literature review**

**Socio-cultural theory views on collaborative learning**

Vygotsky suggests that individuals should be exposed to tasks which are beyond their existing knowledge and which they are unable to solve independently. Accordingly, by interacting and giving assistance with/from others, “imitation” is possible (Chaiklin, 2003). Additionally, in order to understand human learning and thinking, Bonk and Cunningham (1998) asserted that one should explore the “context and setting in which that thinking and learning occurs.”

Socio-cultural theory views learning in a social context (Jeon, 2000). It describes a situation where social and individual processes are interdependent to jointly achieve knowledge construction (John-Steiner & Mahn, 1996). Nevgi, Virtanen and Niemi (2006) indicated that social interdependence variables (e.g., individualistic aspects) are highly correlated to group processes. According to John-Steiner and Mahn (1996), original works of socio-cultural importance are by Vygotsky. This author explored the idea of human activities being mediated by language and other symbol systems. While taking place in a cultural context, human activities are best explained with reference to their historical development.

Human participation in various joint activities will be exposed to wide opportunities for learning from each other, and thus forming the basis for cognitive and linguistic mastery (John-Steiner & Mahn, 1996). As such the focus of socio-cultural research is about the processes for internalization, appropriation, transmission or transformation in formal and informal settings for knowledge to be finally co-constructed (John-Steiner & Mahn, 1996).

With respect to social interaction that is central to collaborative learning, collaborative learning builds upon the socio-cultural theory where a causal relationship exists between social interaction and individual cognitive change (Dillenbourg et al., 1996). Both are interdependent for the co-construction of knowledge (John-Steiner & Mahn, 1996).

Collaborative learning is the term used to describe the learning approach aimed at constructing knowledge to solve a given task by a group of individuals of diverse cultural backgrounds (Alavi & Dufner, 2005; Crook, 1998). Collaborative learning provides opportunities to explore socio-cultural issues to a greater extent with the understanding that collaboration is a philosophy of interaction that requires individuals to be responsible for their actions where besides learning, they are committed to appreciating their friends’ contributions (Panitz, 1996).

As much as socio-cultural theory highlights interaction, the value of interaction in collaborative learning is also denoted by Gerlach (1994) where collaborative learning is a naturally social act as the participants talk among themselves. Collaborative learning considers working together will result in better understanding as compared to working independently (Panitz, 1996).

**Collaborative learning discussions**

According to Dillenbourg (1999), the perception that collaborative learning consists of two or more people learning together does not necessarily describe collaborative learning. This is because ‘two or more people’ can be a group of people of up to 30 members, ‘learning something’ can be simply going through courses or studying specific subjects, and ‘together’ can be learning face-to-face or computer-mediated (Dillenbourg, 1999). However, recent literature has been found to portray collaborative learning in a more explicit way with the description that collaborative learning should be:
First of all, collaborative learning should consists of a small group of students working together. Barkley, Cross, and Major (2005) specifically mentions that groups in collaborative learning should be small, as small as in pairs. Göl and Nafalski (2007) added that a group should be small, not exceeding five members in each group. However, earlier, Dillenbourg (1999) disputed that group size, group heterogeneity and group compositions are variables that interact with each other. Different group identities have different effects on different tasks. What can be done is to explore the influence of the group’s properties towards learning (Dillenbourg et al., 1996).

Also, although interacting with peers can achieve results in learning, they may not learn anything on the basis that they have nothing to work on (Dillenbourg et al., 1996). Thus, there should be specific activities that peers have to perform jointly for learning. As indicated earlier, the term ‘learning’ can refer to following courses or undertaking specific subjects. However, joint problem-solving is the more appropriate activity for collaboration (Dillenbourg et al., 1996). Problem solving will result in members of the group having to share ideas, consider the opinions of others’ and give and receive help (Blumenfeld et al., 1996). Along the process, students will be exposed to more complex skills of planning, monitoring and evaluating progress (Blumenfeld et al., 1996). Problem solving has been chosen as the type of task for collaborative learning in numerous studies (see works by Chung et al., 2008; Puntambekar, 2006; Beers et al., 2005; Cheung, Tan & Hung, 2004).

The ultimate goal for collaborative learning resides within the co-constructed knowledge as a result of learning (Dillenbourg & Fischer, 2007). Upon sharing information among group members and reaching a common ground, knowledge is constructed and later internalized at an individual level (Beers et al., 2005). Collaborative learning values the processes that took place during learning, which is before a specific learning outcome is achieved. It was perceived that collaborative learning, as being central to socio cultural theory, is a result of interaction between individuals (Vygotsky, 1978).

Central to collaborative learning is knowledge construction where the collaborative learning approach aims at co-constructing knowledge upon sharing information in groups for solving given tasks (Dillenbourg & Fischer, 2007; Alavi & Dufner, 2005; Crook, 1998). Knowledge construction is vivid evidence of collaborative learning taking place (Alavi & Dufner, 2005).

In Malaysia, students construct knowledge using strategies such as questioning, clarifying or giving support (Mansor & Abd Rahim, 2009). However, students' behaviour of constructing knowledge is dependent on the online platform being used. For example, when using Facebook as the medium of constructing knowledge, 47.1 percent of the students were found to be able to clarify knowledge by elaborating on ideas and thoughts (Idris & Ghani, 2012). However, when using online discussion boards, such as in Quickplace, students were found to be able to construct knowledge but were limited to seeking and giving opinion as they preferred to impart knowledge. There is no indication of new knowledge being constructed (Hong & Lee, 2008).

The following discussion will provide a better insight on how collaborative learning will engage learning through social knowledge construction.

**Students’ knowledge construction**

Knowledge construction (KC) represents one of the processes for indicating cognitive activities. It can take the forms of, namely: the provision of students’ behaviour of seeking, interpreting, analyzing and summarizing information, critiquing and reasoning through various options and arguments and making decisions in online discussions (Zhu, 2006).

Van der Meijden (2005) affirmed this type of provision for cognitive activities by analysing students’ elaborations in synchronous and asynchronous discussions. She stated that students’ KC is indicated by their behaviour of “posing of comprehension questions requiring explanations, the supply of answers with arguments or justification, the
presentation of new ideas accompanied by explanation, and the acceptance or rejection of ideas coming from others accompanied by arguments” (Van der Meijden, 2005).

In the context of CSCL, knowledge construction is perceived as the eventual outcome of learning. However, the processes of construction and reconstruction are complex and are influenced by interconnected factors (Onrubia & Engel, 2009). This stance was shared by Beers et al., (2005) who indicated that, for unshared knowledge to be finally constructed, intermediate processes such as externalization, internalization, negotiation and integration take place.

Students with diverse expertise need to reach common ground for knowledge to be constructed (Puntambekar, 2006). Within this intermediate process lies the fundamental idea of collaborative learning on how divergent ideas come to work together for knowledge construction (Puntambekar, 2006). Over this concern, Schellens and Valcke (2006) attempted to foster knowledge construction among university students through asynchronous discussion groups using the information processing approach to learning. They theorized that for knowledge to be constructed, an individual’s cognitive processing will be actively engaged. Higher knowledge construction is operationalized as the higher level explicitation type of comment and higher level evaluation type of comment. They found that students in discussion groups are very task-oriented, where higher proportions of advanced phases of knowledge construction are observed (Schellens & Valcke, 2006). Previous discussions demonstrated that students’ knowledge construction can be fundamentally explored from the way they expressed themselves in written communication or discourse.

Analysing students’ knowledge construction strategies in CSCL groups

There are various methods applied by the researchers to understand students’ knowledge construction in CSCL discussions. The common method would be content analysis; where by assigning unit of analysis to the posted messages, the messages can be coded using specific coding scheme and conclusion can be drawn from the findings. For example, Van der Meijden (2005) explored students’ knowledge construction during CSCL discussions and categorized the students into having high or low level of knowledge construction. Although students were discussing in groups, this study did not explore on how the students work as a group to strategize for knowledge construction. A recent study by Lan et al., (2012) explored students’ knowledge construction and their behavioural patterns when discussing in groups using content analysis and sequential analysis. They explored students’ knowledge construction on individual basis but did not compare how a specific group put up strategies to construct knowledge. Other similar studies used sequential analysis to investigate knowledge construction upon using specific tool for learning (see Lin et al., (2012)). However, the findings of these studies were dependent on the tools that they developed to trigger knowledge construction and did not focus on how CSCL group functions.

Research on better understanding on how the students strategize to construct knowledge particularly when discussing; in collaborative group is sparse. There is a need for understanding how group functions during CSCL discussions; that is to learn how the group apply strategies to construct knowledge. Mazzolini and Maddison (2007) pointed that the educators have difficulties in identifying the correct moment to intervene in order to improve the quality of online discussions. As such, this study will serve as an important indicator for educators in conducting online learning discussions particularly when using collaborative learning as the instructional strategy.

Method

Participants

The participants were 20 students enrolled in the subject Web-based Multimedia Development at a university in southern Malaysia. There were five groups consist of four members in each group. All the students were familiar with the online discussion environment because they have used the same platform for other subjects as well.

CSCL problem-solving tasks for discussion

Students in the course have to solve five problems in groups related to the Web-based Multimedia Development subject. The problems were real-life problems that the students have to face whenever designing educational
websites as related to this course. Students were given two weeks to solve each problem through group discussions in the course discussion board. Students were not assigned to any particular role in groups to maintain collaboration to occur as natural as possible. The instructor gave scaffold to group discussions such as inviting participation and giving clues. The students solved each of the problems in their own group in different discussion section. The following are the tasks that all the students have to solve in their groups in order:

- Week 2 - Problem: Error 404,
- Week 4 - Problem: Problems in website design,
- Week 6 - Problem: Pictures in a website: why my picture did not appear in web browser?
- Week 8 - Problem: Different views in Opera, Mozilla, and Google Chrome: What happen to my website?
- Week 10 - Problem: Which media fits in website? Which media does not fit?

Research design and procedures

This research adopted a case study research design to answer questions on how students’ construct knowledge during CSCL discussions. Next, comparison is made to the CSCL groups’ strategy in constructing knowledge. The content analysis and sequential analysis technique were used to finally construct transition state diagrams of the CSCL groups.

Content analysis procedures

To investigate students’ behaviour of constructing knowledge in CSCL discussions, this study used content analysis technique to analyse students’ discussion scripts. Content analysis is applied to students’ discussion scripts while discussing in online learning discussion board based on a coding scheme. Upon characterization based on the coding scheme, this study applied unit of meanings for objectivity. This research adopted the coding scheme by Van der Meijden (2005) who developed the coding scheme to evaluate students’ knowledge construction in asynchronous and synchronous online discussions as in Table 1.

Table 1. The coding scheme for analysing students’ knowledge construction in CSCL discussions

<table>
<thead>
<tr>
<th>Cognitive: Asking questions (Cognitive 1)</th>
<th>Examples of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHV 1 Asking questions that do not require an explanation (facts or simple questions)</td>
<td>Has the problem been solved?</td>
</tr>
<tr>
<td>*CHV 2 Asking questions that require an explanation (comprehension or elaboration)</td>
<td>You have explained all the units for developing the website, but which one is preferable and why?</td>
</tr>
<tr>
<td>CHVER Verification or asking for agreement</td>
<td>Is it true?</td>
</tr>
<tr>
<td></td>
<td>Am I explaining correctly?</td>
</tr>
<tr>
<td>Cognitive: Giving answers (Cognitive 2)</td>
<td></td>
</tr>
<tr>
<td>CHG 1 Answering without explanation</td>
<td>There are 3 types of images.</td>
</tr>
<tr>
<td>*CHG 2 Answering with explanation (using arguments or by asking a counter-question)</td>
<td>The problem has been solved.</td>
</tr>
<tr>
<td></td>
<td>.jpeg is different than .png image.</td>
</tr>
<tr>
<td></td>
<td>It means that not all computer resolution is the same because..</td>
</tr>
<tr>
<td></td>
<td>The information shows that ..</td>
</tr>
<tr>
<td>Cognitive: Giving information (Cognitive 3)</td>
<td></td>
</tr>
<tr>
<td>CI 1 Giving information (an idea or thought) without elaboration</td>
<td>I paste the information from the internet as presented below ..</td>
</tr>
<tr>
<td>*CI 2 Giving information (an idea or thought) with elaboration</td>
<td>From what I see, both images look the same.</td>
</tr>
<tr>
<td></td>
<td>I guess the alternative way to solve this is by ..</td>
</tr>
<tr>
<td></td>
<td>From the example that I obtained from the internet below, they say that ..</td>
</tr>
</tbody>
</table>
For reliability purposes, two raters coded 55 messages in 6 continuous coding sessions. “Message” is the term used to describe the raw discussion scripts posted by the students. In every session, we coded the messages individually and then compared our codes. We discussed, compared and contrasted on our codes to finally reach a common agreement on the suitable code for each meaning in the students’ messages. The processes were repeated until we finally gathered all the messages and coded them in the final coding session.

In the final coding session, the number of unit of meanings was calculated to clarify that both coders had the same number of segments. We used the term ‘segment’ to signify discussions scripts with specific unit of meaning. There were a total of 105 segments from the 55 messages.

Next, we compared the codes on the segments to look for agreement or disagreement. From 105 segments, it was found that both coders agreed upon 89 segments which accumulated 84.76 percent of agreement. Coding session resulted in 84.76 percent of inter-coder agreement which is higher than 83 percent as obtained by Paulus (2009) and it is higher than 81.4 percent (Van der Meijden, 2005) when inter-rater reliability was calculated; and thus acceptable. The resulting kappa value was 0.745 where Fleiss (1981) and Capozzoli et al. (1999) stated that a Kappa value of 0.40 – 0.75 as intermediate to good. This indicates that both coders have good mutual understanding about applying the codes to students’ discussion scripts.

Finally, students were identified to be either in the group with high level of knowledge construction (H), high-low level of knowledge construction (HL) or low level of knowledge construction (L). The identification of groups is summarized in Table 2. We compared the frequency of codes of high-level cognitive contribution (as indicated with “*” in Table 1) with the frequency of low-level cognitive contribution. Groups that posted more high-level codes than low-level codes are categorized as “H” group, groups with more low-level codes than high-level codes are categorized as “L” group and groups with same frequency of codes in high and low-level are categorized as “HL” groups.

<table>
<thead>
<tr>
<th>Details</th>
<th>Group type</th>
</tr>
</thead>
<tbody>
<tr>
<td># High-level cognitive contribution &gt; # Low level cognitive contribution</td>
<td>H</td>
</tr>
<tr>
<td># High-level cognitive contribution = # Low level cognitive contribution</td>
<td>HL</td>
</tr>
<tr>
<td># High-level cognitive contribution &lt; # Low level cognitive contribution</td>
<td>L</td>
</tr>
</tbody>
</table>

**Table 2. Group identification based on level of knowledge construction**

**Sequential analysis procedures**

To further understand students’ strategies in constructing knowledge, this research used sequential analysis technique. Sequential analysis technique shows the sequences of students’ action when constructing knowledge. The coded 292 segments from content analysis underwent a series of sequential transition matrix calculation (Bakeman & Gottman, 1997). This research considers a z-score greater than 1.96 as the sequence of a row and a column as being statistically significant ($p < 0.05$) (Bakeman & Gottman, 1997). Finally, the significant sequences were gathered to
construct the transition state diagram that illustrates the different students’ strategies to construct knowledge in CSCL discussions for groups with different levels of knowledge construction.

Results

This study investigates students’ behaviour when constructing knowledge during problem solving discussions using content analysis and sequential analysis techniques. The percentages of the coded segments from content analysis procedures are shown in Figure 1. There was no student who posted messages from the code CHV 1 and NAN thus the result is not displayed in the figure.

Generally, students posted most messages at the low-level of knowledge construction. Highest percentage is accumulated at the CI 1 code (25.81%) which contributed almost half of the percentage in low-level of knowledge construction. The CI 1 code was most demonstrated when students in this research “paste” the information directly from internet resources without giving elaboration on the information during online discussion. However, some students made summary from the information that they shared. For example, Student S16 posted the following message:

Student S16: Oh my friends, lets’ check here:
Many people ask “Why is my website design different when using Mozilla and IE?” How do I get .. (CI 1).

Summary: It is said (from the above information) that this problem exists because of the validation process that used HTML and XHTML version (CIE).

In this research, it is also found that the higher-level of knowledge construction such as CHV 2, CHG 2, AY and NAY are difficult to reach due to low percentages. However, looking at the bright side, there are in fact argumentations (AY and NAY) going on and active question and answer sessions (CHV 2 and CHG 2) although at low percentages. For example, Student S2 posted the following:

Student S2: I’m quite confused as to which one is better for a website. Can anyone tell me the quality difference between .gif and .swf format? I try to search for .. (CHV 2).

With the exposure to CSCL discussions, students in this study were able to display behaviors that reflected high-level of knowledge construction such as “asking questions that require an explanation (comprehension or elaboration)” (CHV 2), “answering with explanation (using arguments or by asking a counter-question)” (CHG 2), “sharing information with elaboration” (CI 2), “accepting information with elaboration” (AY), and “not accepting information by providing elaboration” (NAY). Students’ pattern of contributions in collaborative groups was identified as in Table 3.
From Table 3, there were three groups identified as having high level of knowledge construction (H), that is, more than half of the groups. There is one group categorized as the “L” group. To further understand how these groups construct knowledge, sequential analysis was carried on the respective groups. Table 4 presents an example of the z-scores obtained from the “H” groups. From Table 4, each row represents a starting knowledge construction strategy and each column represents the follow-up strategy. The significant sequence is the sequence with z-score of more than 1.96 (Bakeman & Gottman, 1997).

Table 3. Group categorization based on levels of knowledge construction

<table>
<thead>
<tr>
<th>Group</th>
<th>#Low-level cognitive contribution</th>
<th>#High-level cognitive contribution</th>
<th>Group type</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>51</td>
<td>17</td>
<td>L</td>
</tr>
<tr>
<td>B</td>
<td>16</td>
<td>21</td>
<td>H</td>
</tr>
<tr>
<td>C</td>
<td>41</td>
<td>50</td>
<td>H</td>
</tr>
<tr>
<td>D</td>
<td>19</td>
<td>27</td>
<td>H</td>
</tr>
<tr>
<td>E</td>
<td>25</td>
<td>25</td>
<td>HL</td>
</tr>
<tr>
<td>Total</td>
<td>152</td>
<td>140</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Example of z-score for the ‘H’ groups

<table>
<thead>
<tr>
<th></th>
<th>CHV2</th>
<th>CHVER</th>
<th>CHG1</th>
<th>CHG2</th>
<th>CI1</th>
<th>CI2</th>
<th>CIT</th>
<th>CIE</th>
<th>AN</th>
<th>AY</th>
<th>NAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHV2</td>
<td>-0.19</td>
<td>-0.43</td>
<td>1.06</td>
<td>0.40</td>
<td>-1.51</td>
<td>2.34</td>
<td>0.71</td>
<td>-0.82</td>
<td>-0.082</td>
<td>-1.04</td>
<td>-0.43</td>
</tr>
<tr>
<td>CHVER</td>
<td>-0.51</td>
<td>-0.19</td>
<td>-0.47</td>
<td>0.80</td>
<td>0.11</td>
<td>-0.69</td>
<td>-0.53</td>
<td>-0.35</td>
<td>-0.35</td>
<td>1.96</td>
<td>-0.19</td>
</tr>
<tr>
<td>CHG1</td>
<td>0.06</td>
<td>-0.38</td>
<td>-0.97</td>
<td>1.64</td>
<td>0.87</td>
<td>-1.42</td>
<td>-1.09</td>
<td>-0.73</td>
<td>2.34</td>
<td>-0.92</td>
<td>-0.38</td>
</tr>
<tr>
<td>CHG2</td>
<td>-0.16</td>
<td>1.33</td>
<td>0.08</td>
<td>0.79</td>
<td>-0.69</td>
<td>0.72</td>
<td>-1.02</td>
<td>0.94</td>
<td>-1.17</td>
<td>0.22</td>
<td>-0.62</td>
</tr>
<tr>
<td>CI1</td>
<td>-0.26</td>
<td>-0.97</td>
<td>1.42</td>
<td>0.36</td>
<td>0.67</td>
<td>-0.75</td>
<td>-1.59</td>
<td>1.48</td>
<td>-0.18</td>
<td>-0.32</td>
<td>-0.97</td>
</tr>
<tr>
<td>CI2</td>
<td>-0.68</td>
<td>-0.55</td>
<td>-1.39</td>
<td>-1.67</td>
<td>-1.95</td>
<td>1.23</td>
<td>3.25</td>
<td>1.04</td>
<td>1.25</td>
<td>2.38</td>
<td>1.57</td>
</tr>
<tr>
<td>CIT</td>
<td>-0.26</td>
<td>2.03</td>
<td>0.95</td>
<td>-0.44</td>
<td>1.21</td>
<td>-0.90</td>
<td>-0.34</td>
<td>-0.85</td>
<td>-0.85</td>
<td>0.01</td>
<td>-0.45</td>
</tr>
<tr>
<td>CIE</td>
<td>0.81</td>
<td>-0.27</td>
<td>-0.67</td>
<td>-1.10</td>
<td>0.15</td>
<td>-0.99</td>
<td>0.73</td>
<td>1.63</td>
<td>1.63</td>
<td>-0.64</td>
<td>-0.27</td>
</tr>
<tr>
<td>AN</td>
<td>-0.66</td>
<td>-0.24</td>
<td>-0.61</td>
<td>-1.00</td>
<td>2.44</td>
<td>-0.90</td>
<td>0.94</td>
<td>-0.46</td>
<td>-0.46</td>
<td>-0.58</td>
<td>-0.24</td>
</tr>
<tr>
<td>AY</td>
<td>1.30</td>
<td>-0.37</td>
<td>-0.92</td>
<td>0.17</td>
<td>-0.24</td>
<td>0.46</td>
<td>0.08</td>
<td>-0.69</td>
<td>-0.69</td>
<td>-0.88</td>
<td>2.58</td>
</tr>
<tr>
<td>NAY</td>
<td>2.22</td>
<td>-0.15</td>
<td>-0.38</td>
<td>-0.63</td>
<td>0.60</td>
<td>-0.56</td>
<td>-0.43</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.37</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

Table 5 presents the summary of significant sequence for “H” groups, “HL” group and “L” group respectively at p < 0.05.

Table 5. The significant sequences according to group type

<table>
<thead>
<tr>
<th>Group type</th>
<th>Significant sequence</th>
<th>z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>High level knowledge construction (H)</td>
<td>CHV 2 -&gt; CI 2</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>CHG 1 -&gt; AN</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>CI 2 -&gt; AY</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>CIT -&gt; CHVER</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>AN -&gt; CI 1</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>AY -&gt; NAY</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td>NAY -&gt; CHV 2</td>
<td>2.22</td>
</tr>
<tr>
<td>High-low level of knowledge construction (HL)</td>
<td>CHG 2 -&gt; CHVER</td>
<td>2.06</td>
</tr>
<tr>
<td></td>
<td>CI 2 -&gt; CIT</td>
<td>3.08</td>
</tr>
<tr>
<td></td>
<td>AY -&gt; CHV 2</td>
<td>3.89</td>
</tr>
<tr>
<td>Low level knowledge construction (L)</td>
<td>CHV 2 -&gt; CHG 1</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>CHV 2 -&gt; CHG 2</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>CI 2 -&gt; AN</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>CIE -&gt; CHV 2</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>CIE -&gt; AY</td>
<td>2.69</td>
</tr>
</tbody>
</table>

Students in different type of group were found to apply different strategies to construct knowledge during online discussions. Comparison was made between the transition state of group “H”, “HL”, and “L” as presented in Figure 2, Figure 3 and Figure 4 respectively. There were seven significant sequences for the ‘H’ groups with most
sequences involving the code AN and AY; the codes for argumentation. In “H” groups, more significant sequences evolved in Cognitive 3 (Giving Information). There were more inter-sequences in Cognitive 3 for “H” groups as compared to the other groups (see Figure 2). Also, it is interesting to note that the “H” groups have argumentation elements (NAY, AY and AN codes) going on as compared to the other groups. For the “HL” group, there were three significant sequences where the z-scores for CI 2 – CIT and AY – CHV 2 were very high, indicating that the events were not by chance. Such strategy, although proven to be significant, is likely to be insufficient to sustain knowledge construction particularly at high-level.

**Figure 2. The transition state diagram for “H” groups type**

**Figure 3. The transition state diagram for “HL” group type**

**Figure 4. The transition state diagram for “L” group type**

In “L” group, although the frequency of high-level codes were low as compared to the low-level codes, there were five significant sequences involving high-level cognitive contributions such as CHV 2, CHG 2 and AY. In terms of knowledge construction, the results showed that the “L” group lacked the strategies of negotiation of information (AY - NAY) as compared to the “H” groups. Although some of the questions posted by the members were of high level (CHV 2), the other members proceed with answering but they did not elaborate on their answers (CHV 2 – CHG 1) (see Figure 4). Such strategy is insufficient for construction of knowledge since they did not have detailed discussions on task.

**Discussions**

This study has revealed several implications for educational practice.
Using argumentation for knowledge construction sustainability

This study found that different groups showed significant different strategies when constructing knowledge. The “L” group focused more on answering questions and the strategy ended at that point (Cognitive 2). Although there were strategies that involved giving detailed information, they lacked the argumentation elements (AY, AN and NAY) even though they did agree on statements and elaborate on them. Agreeing on statements seems insufficient for the group to reach high level of knowledge construction as a group. According to Howard (1996), by arguing, justifying, explaining and providing evidence, students are self-regulating and hence, they are communicating at a high level of cognitive engagement. Likewise, argumentation is always needed to learn in conjunction with problem-solving.

The “H” groups have more thorough discussions as compared to the other groups. According to the transition state diagram, they negotiate, compare and contrast before proceed to the next strategy for constructing knowledge. Interestingly, despite other high level codes, the argumentation codes (AY, AN and NAY) were the codes that differentiate “H” groups than the other groups. For example in “H” groups, when they came into argument, they provide justification and give information (sequence AN – CI 1). Accordingly, whenever they disagreed with information, they proceed with asking questions that require elaborations (sequence NAY – CHV 2).

This study suggest that triggering argumentation is very important and it was the strategy that maintained the “H” groups to have high-level of knowledge construction as compared to the other groups. As such, to ensure group functioning, educators could assist discussions to promote argumentation such as by asking questions or to provoke opinions during the discussions.

Problem-solving tasks for CSCL discussion

This study found that problem solving discussion can influence students’ behaviour of constructing knowledge in groups. Although this study showed that the highest frequency during online discussion is CI 1, all the groups (“H,” “HL,” and “L” groups) were found to be able to apply strategies that involved high-level of knowledge construction codes such as AY and CI 2. This might due to the authentic problem solving task given to the students that requires them to spend extra effort to successfully solve the problems. Collaborative problem solving will exposed students to various solutions and they have to tackle different opinions. They are also able to determine their own attitude while dealing with problems (Rahikainen, 2002). It is similar to results reported by Hurme and Järvelä (2001) that problem solving is the appropriate tasks for collaboration because it helps to reveal the metacognitive processes as an important output of learning. However, problem solving tasks for CSCL discussion has to be properly design so that it will trigger argument. In this study, argumentation is found to play one of the primary roles towards students discussing at the high level of knowledge construction. Possibly, organizing debates would yield more groups functioning at the high level of knowledge construction.

Limitations and future research recommendation

This study has several limitations. Firstly, this study is interested in discovering how students construct knowledge while discussing in CSCL. Thus, the results of this research only came from the data obtained from online discussion for in-depth understanding. Categorization of students’ knowledge construction in groups also relies heavily on their contributions in online learning discussions. The results of other students’ activities in CSCL is has not taken into account in the research. However, it is important to know that knowledge construction is difficult to create and sustain, thus exploring the facets that could trigger knowledge construction can be considered a success. Other than that, this study did not consider the motivational aspects and the group composition (gender) that might also influence the group strategies when constructing knowledge. This study is limited to understanding how CSCL group can function when constructing knowledge. Further research should explore in-depth on the influences of different variables towards students’ strategy of constructing knowledge particularly when solving problems.

Conclusions

Conclusively, this research used content analysis and sequential analysis to explore students’ strategies when constructing knowledge in online discussions. After the analysis, students’ knowledge construction were explored
with respect to their CSCL groups and transition state diagrams were drawn according to the group’s type of either “H,” “HL,” and “L.” It was found that students still struggle to construct knowledge that involved elaborations. There were significant differences in application of strategies to construct knowledge in groups. The “H” type of groups focussed more on negotiation of information by arguing and debating while the “L” group was comfortable with asking-answering sessions. We also discovered that for more comprehensive knowledge construction during CSCL discussions, argumentation is one of the crucial factors. As such, the problem solving tasks for CSCL should trigger argumentation elements such as arguing, justifying, explaining and providing evidence.

Acknowledgements

The authors would like to thank the Universiti Teknologi Malaysia (UTM) and Ministry of Education (MoE) Malaysia for their support in making this project possible. This work was supported by the Research University Grant (00K61) initiated by UTM and MoE.

References


The Effect of Using Cooperative and Individual Weblog to Enhance Writing Performance

H. Gulhan Orhan Karsak1*, Seval Fer2 and Feza Orhan3

1Social Sciences Institute, Yıldız Technical University, Turkey // 2Department of Education Programs and Instructional Design, Hacettepe University, Turkey // 3Department of Computer Education and Instructional Technology, Yıldız Technical University, Turkey // gorhan811@gmail.com // sevalfer@hotmail.com // feza.orhan@gmail.com

Corresponding author

(Submitted April 25, 2013; Revised November 09, 2013; Accepted January 08, 2014)

ABSTRACT

Academic writing, whether individual or cooperative, is an essential skill for today’s graduates. However, motivating and helping students to learn to write effectively, either in cooperative or individual scenarios, poses many challenges, many of which can be overcome by technical means. The aim of this study is to investigate the effect of using cooperative and individual weblogs to enhance writing performance, based on blended instructional design. An equally weighted sequential quantitative-qualitative (two-stage) mixed research design was used for this study. In the quantitative dimension, the “pretest-posttest control group design” that was applied as an experimental study was used. In the qualitative dimension, the views of students in the experimental group were examined as a “case study.” Twenty-one 5th grade students participated in the study, 12 in the experimental group, and 9 in the control group. While the experimental group worked on a cooperative weblog, the control group worked on an individual weblog. A writing performance exam was applied as pretest-posttest. Exam papers were scored with a writing performance evaluation scale. Results of the quantitative dimension indicated a statistically significant difference between the writing performance of students in the experimental and control groups in favor of the control group. When the sub-skills of students’ writing performance was examined on the posttest scores of the “ideas & content dimension,” a statistically significant difference was found in favor of the control group, but no significant difference was detected in the “sentence fluency, writing rules and organization” dimensions between the groups. Results of qualitative dimension showed that “feedback” has a positive effect on writing performance, unlike cooperative learning, which has an adverse effect.

Keywords

Blended instructional design, Cooperative learning, Weblog, Writing performance

Introduction

According to Graham and Harris (2009), many students show limited writing performance due to the fact that they do not make a good writing plan, have difficulties in creating content, do not know how to use writing strategies, their revision of the writing process is not meaningful enough, and they do not make enough effort to write. Obviously, writing enters into all aspects of human life, especially nowadays in the age of the Internet and email.

Vincent’s (2003) research in which an 11 year old student, who does not like to study writing activities and cannot write more than 1-2 lines of work in the paper-pencil studies, also supports this idea. In accordance with this argument, in her research that writing exercises through weblogs and social networks conducted by 3001 students aged between 12-13 years from 12 elementary and 12 middle schools, Clark (2009) stated that youngsters cannot make compound sentences, do not have a wide range of vocabulary or use a variety of words in their writings, and furthermore, they do not pay attention to capitalization, punctuation, and spelling.

In order to improve writing performance, students should observe the environment dedicatedly, and produce ideas on what they have observed and read. Then they must show their improvements gained through these means in the appropriate learning environments with the support of an effective teaching plan enhanced by various techniques (Chaffee, McMahon and Stout, 2004).

It was notable in the literature that cooperative learning is suggested to teach writing effectively. For instance, Johnson, Johnson and Stanne (2000) expressed that cooperative learning enables students to work together for both
their own learning and the other group members’ learning. It also enables each student to take responsibility for group members’ learning. Santangelo and Olinghouse (2009) and Graham and Perrin (2007) also state that well-constructed cooperative writing activities help students learn from each other and share their cultural backgrounds. In addition, it increases the quality of writing activities. Similarly, Santangelo, Harris and Graham (2007) indicate that cooperative writing activities give students the opportunity to take an active role in planning their writing with their peers and teachers through positive interaction, writing, revising, and editing.

Developing cooperative writing activities that have the abovementioned characteristics in an enjoyable learning environment that attracts students’ attention and enables them to socially interact with their peers and teachers without constraints on time or place is very important to increase writing performance. An online learning environment is the first environment that comes to mind in accordance with the qualities indicated above. Davidson-Shivers and Rasmussen (2006) define an online learning environment as an environment that any type of tasks can be given and in which students interact with each other and their teachers without face-to-face interaction, as in the classroom. In order to overcome this limitation, conventional face-to-face learning and online learning can be combined to create an effective learning environment for writing activities. That is to say that the strong interaction and communication opportunities that face-to-face learning offers and online learning’s advantage of being independent of time and setting have fostered the idea of developing an environment for teaching writing based on blended learning, which combines face-to-face learning and online learning.

A literature review revealed that blended learning is generally defined as the integration of internet-based learning and face-to-face learning (Bonk and Graham, 2006; Kerres and Witt, 2003; Rooney, 2003). In an effective blended learning environment, learning tools, which include messenger, tele-video conferencing, debate rooms, forums, emails, and web 2.0 tools, are as important as teachers (Cuhadar, 2008).

Weblogs or blogs are one of the prominent Web 2.0 tools in the interactive web technology environment. Blogs are defined by different writers (Babbage and Wasson, 2006; Glogoff, 2003; Martindale and Wiley, 2005) as web environments that are easy to create and through which links such texts, pictures, and audio and video files can be shared and updated; additionally, individuals from different locations at different times can comment on these links.

Using blogs in education, in addition to the advantage of archiving, provides teachers and students the opportunity to interact and share their thoughts with each other and with other people. Furthermore, blogs increase students’ individual responsibility by publishing their performance and aid in their entertainment by reading the comments on their performance and enabling them to think critically (Du and Wagner, 2007; Shelly, Gunter and Gunter, 2010). Another useful aspect of blogs in learning, as Farmer and Barlett-Bragg (2005) stated, is that students who tend to learn both individually and in a group have reading and writing experiences by internalizing the blog environment. When the body of literature is observed, it is seen that except diary writing, the blog writing study conducted by Blankenship, 2007; Huffaker, 2005; Kelly and Safford, 2009).

Accordingly, Graham and Harris (2009) stated that by taking into consideration the students’ common facilities, interest areas, and individual differences, an “individual weblog environment” based on blended instruction design can systematically bring the deficient characteristics (making a good plan, using an appropriate environment and strategy, sufficient application, and meaningful review) that are necessary for good writing performance. Chen et al. (2011) also supports this idea with the results of the blog writing study conducted on 5th grade students over 15 weeks. Their study results indicated that students gained the skills of writing meaningful and rich content, organization, fast writing, and gathering information via the internet.

**Purpose of study**

In the light of abovementioned research results and discussions, this study aimed to examine the effect of cooperative and individual weblog-integrated writing instructions, based on blended instructional design, on 5th grade students’ writing performance. The expectation was to obtain answers to the two research questions below.
• Is there a statistically significant difference between the writing performance of students who were in cooperative weblog-integrated learning environment and who were in individual weblog-integrated learning environment which are based on blended instructional design?

• What are students’ opinions about the effects of cooperative weblog-integrated writing instruction on their writing performance?

Method

Research design

An equally weighted sequential quantitative-qualitative (two-stage) mixed research design was adopted for this study.

For the quantitative aspect, the effect of the two independent variables, cooperative and individual weblog-integrated writing instructions, based on blended instructional design, on writing performance, which is the dependent variable of the research, was examined. For this reason, in this aspect, “pre- and post-test control group design,” which is one of the experimental designs, was applied. Experimental design shows in Table 1.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pre-test</th>
<th>Experimental process</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group I</td>
<td>O₁₁</td>
<td>X₁</td>
<td>O₁₂</td>
</tr>
<tr>
<td>Experimental Group II</td>
<td>O₂₁</td>
<td>X₂</td>
<td>O₂₂</td>
</tr>
<tr>
<td>Control Group</td>
<td>O₃₁</td>
<td>X₃</td>
<td>O₃₂</td>
</tr>
</tbody>
</table>

In Table 1, “O” shows the implementation of the writing performance exam, which consists of a pre-test and post-test; “X,” shows the implementation of writing performance in an “individual weblog environment”; X₂ shows the implementation of cooperative weblog-integrated writing instruction environment based on blended instructional design in experimental groups; and X₃ shows the implementation of face-to-face writing instruction in control groups.

For the qualitative aspect, the case study method was used to reveal opinions of the students in the experimental group about the effects of cooperative weblog-integrated writing instruction, on their writing performance.

Study groups and equivalency process

The study group consisted of students from two classes of an elementary state school in Istanbul, which possessed all the required physical features and was convenient for the researchers in the spring term of 2010-2011 academic year.

Study groups

The experimental process was conducted with 21 students who were divided into experimental and control groups. Students in the experimental group (n = 12) were randomly chosen from the class 5/B, and students in the control group (n = 9) were randomly chosen from the class 5/C.

In the experimental group, there were 7 (58.3%) female and 5 (55.6%) male students, and in the control group there were 3 (33.3%) female and 6 (66.7%) male students.

Equivalency process

Before the experimental process, a “Writing Performance Test,” which is described under the title “Data Collection Instruments,” was applied to the entire study group as a pre-test, to create equivalency between the experimental group and the control group in terms of their writing performance. After the pre-test, the Mann-Whitney U-test was used to determine the equivalence between the experimental and control groups. Table 2 shows the values for pre-test scores of the groups.
As is seen in Table 2, there is no significant difference between pre-test scores of the experimental and control groups in terms of ideas/content (U = 42.00, p > .05), coherence (U = 45.50, p > .05), organization (U = 46.50, p > .05), and writing rules (U = 48.50, p > .05) aspects and of their writing performance (U = 44.00, p > .05). It can be inferred from this that the groups have equal basic writing skills.

Qualitative aspect of the research

It was conducted by 12 students, including 7 females and 5 males, from experimental group.

Data collection instruments

In order to answer the research questions, data was collected through the three instruments explained below:

Writing performance test

The scale that was used as a pre- and post-test to measure the writing performance of the students was developed by Seval Fer and H. Gulhan Orhan Karsak. In this test, the open-ended question directed to students was, “Think of all the branches of fine arts and their characteristics and write a story including either only one branch or all.” Answer sheets were evaluated by three different primary school teachers, without any interaction, using a writing performance evaluation scale.

Writing performance evaluation scale

The writing performance evaluation scale was developed to evaluate the writing performance of the students.

The scale, which was originally developed by Seval Fer and H.Gulhan Orhan Karsak for this research, consists of 25 items testing basic writing performance, including ideas/content (13 items), organization (3 items), fluency (6 items), and writing rules (3 items). Each item is graded in four categories as: very good (4), good (3), adequate (2), and inadequate (1). According to the evaluation of eight experts of the field, no editing was made on the scale. The consistency among the evaluations of three 5th grade teachers based on the test scores, which are used to determine the writing performance of the students, was examined through the Pearson Correlation Analysis. As the basic writing performance scores were analyzed, the highest correlation was detected between the 2nd and the 3rd evaluators (r = .98; p < .05), while the lowest correlation was detected between the 1st and the 2nd evaluators (r = .94; p < .05). The high level of positive and significant correlation among the evaluators indicates that the evaluation was consistent and reliable.
Interview form

For the qualitative aspect of the research, an interview form including two questions was used to identify the students’ opinions on the effect of cooperative weblog-integrated writing instruction based on blended instructional design on their writing performance. Special attention was given to ensure that the interview questions reflected the students’ views on the effect of cooperative weblog-integrated writing instruction based on blended instructional design on their writing performance. The questions were preferred because they were appropriate for focus groups and the possibility for students to add their ideas.

Learning environment

Cooperative weblog-integrated instructional design was applied to the experimental group in a blended learning environment, while individual weblog-integrated instructional design was applied to the control group. Some of the activities were implemented by the teachers of groups over four weeks. During the first week, four-hour preparatory activities were implemented. In the following three weeks, three-hour writing activities were completed per week. Activities in the learning environment were observed by two volunteer observers to determine whether the stages of instructional design were applied during the teaching process or not. One of the observers was the teacher of another class in the same school and the other was one of the researchers.

Working process of the experimental group

Cooperative weblog-integrated instructional design in a blended learning environment, which was developed to be applied in the experimental group, is based on the ADDIE model that has the following main and substages.

- Analysis: In the context of “Learner Analysis,” which is the first step of this stage, a preference form was prepared by referring to the 2005 Ministry of National Education Turkish Curriculum in order to ascertain students’ preferences for writing activities. According to the preferences of most students, story writing was determined as the most preferred genre, and the four most preferred fine arts branches were theater, painting, music, and sculpture, respectively.

A writing test was applied as a pre-test to measure students’ writing skills. In the context of “Determining Tasks and Objectives,” which is the second step of the Analysis stage, after defining the general aims, expected writing tasks related to these aims were determined and divided into four groups: ideas/content, fluency, organization, and writing.

The writing objectives related to the defined writing tasks were taken from the 2005 Ministry of National Education Turkish Curriculum for Primary Education.

- Development: For the “Determining Teaching Strategies and Methods-Preparatory Activities” step, which is the first step of the Development stage, implementer teachers and participant students were provided with information through a power point presentation. Implementer teachers were informed about the instructional design plan and implementation, and the students were informed about the characteristics, processes, and rules of writing a story; fine arts branches that were preferred by students; features and uses of weblogs and about the forms used. Students also received computer literacy training, including punctuation and capitalization. In the second step of the Development stage, “Presenting Tasks,” three cooperative groups were formed, each of which had four students. Each of the students in the cooperative groups participated in one of the sub-themes specified in the reading instruction. In context of “Writing Activities,” the jigsaw technique, one of the cooperative learning techniques, is used, and includes stages as defining learner characteristics, preparing materials, forming groups, distributing the subjects, working in cooperative groups and expert groups, exchanging information in groups, and individual assessment. In the context of “Feedback”, which is the fourth step of the Development stage, the aim was to enable the teacher to guide writing activities. Students wrote comments to each other, as well as to the teacher, who was tasked with writing to students regularly each week by using the evaluation forms.

In the fifth step of the Development stage, “Deciding and Preparing Teaching Materials and Setting” a weblog address was assigned to each student by researchers via the blog server called “Blogcu” (http://www.blogcu.com). For the preparatory activities, PowerPoint presentations were prepared, and for the writing activities, reading
instructions and a draft form for story writing were prepared. In the context of the last step, “Developing Assessment Instruments,” teacher assessment, group work self-assessment, and group work peer assessment instruments were developed using the same items with different names. The pre-test was envisaged to be used as the post-test.

- **Implementation:** In this stage, all the steps of the Development stage were implemented.
- **Evaluation:** The evaluation was completed by using the “Writing Scale,” “Group Work Self-Assessment Form,” “Group Work Peer Assessment Form” and “Teacher Assessment Form” by referring to the writing tasks and objectives that were defined in the second step of the Analysis stage. The post-test was applied.

**Study process of the control group**

The individual weblog-integrated instructional design in a blended learning environment, which was developed to be applied in the control group and based on the ADDIE model, has the following main and sub-stages.

- **Analysis:** The steps followed in this stage are the same as the steps defined in the Analysis stage of the instructional design developed for the experimental group.

- **Development:** The first steps of this stage are the same as the steps defined in the development stage of the instructional design developed for the experimental group. On the other hand, in the context of “Presenting Writing Tasks,” the second step of this stage, each students’ participation in the themes in the reading instruction and write in the blogs was individually enabled. The third step of this stage, “Writing Activities,” allowed students to obtain information about the specified fine arts theme and answer the questions related to the theme by using the websites given in the reading instructions. Next, the students were requested to complete the draft form for story writing. The students were also required to complete their writing tasks individually, depending on the plan, and by considering the writing processes. The “Feedback” step, the fourth step of “Development” stage, the teachers guided the students during the writing activities; enabled the students to write comments to each other and to the teacher who was tasked with writing to the students regularly each week by using the evaluation forms.

“Deciding Teaching Materials and Setting” was the fifth step of this stage, in which it each student was assigned a weblog address by researchers via the blog server called “Blogcu” (http://www.blogcu.com). In preparation of this step, PowerPoint presentations were prepared, and for writing activities, reading instructions and a draft form for story writing were prepared. In the context of the last step of this stage, “Developing Assessment Instruments,” instruments for teacher assessment were developed, as well as group work self-assessments and group work peer assessment. The students were requested to use the assessment instruments individually.

- **Implementation:** In this stage, all the steps of development stage were implemented.

- **Evaluation:** The evaluation forms that were used are the same as the forms used in the Evaluation stage of the instructions designed and developed for the experimental group. After the implementation, each student evaluated his/her own writing performance by completing the “Self-Assessment Form” and evaluated one of the group member’s writing performance by completing the “Peer Assessment Form.”

**Data analysis**

**Quantitative aspect**

The first research question aimed to reveal the effects of cooperative and individual writing activity based on blended instructional design in which weblog and face to face environments exist on students writing performance. Cooperative and individual blended instructional designs are the two independent variables and students’ writing performance is the dependent variable of the research. The Mann-Whitney U-Test was used to answer the research question. The significance level was accepted as \( p < 0.05 \).
Qualitative aspect

In the qualitative aspect of the research, a focus group interview was conducted on 12 students from the experimental group. The questions in the interview form were directed to the students. The responses of the students were recorded by the teacher of the experimental group, while the researcher observed the interview.

At the end of the interview, recorded responses were decoded into text. The content analysis of the data gathered by the focus group interview was done. The content analysis was done by the means of NVivo 8 qualitative data analysis program.

To test the validity and the reliability of the research, the following steps, defined by Yildirim and Simsek (2000, 76), were followed: (1) For external reliability, the researcher remained in the background and participated in the process as an observer. For this purpose, the focus group interview was conducted by the implementer teacher; (2) To provide internal validity, the goal was to obtain internally consistent and significant results; (3) To provide external validity, the study group, setting, and the processes were explained in detail.

Results

Results of the research were examined from quantitative and qualitative aspects.

Results of the quantitative aspect

The first research question was expressed as, “Is there a statistically significant difference between the writing performance of students who received writing instructions in a cooperative weblog-integrated learning environment based on blended instructional design, and students who received writing instructions in an individual weblog-integrated environment?” In this context, the Mann-Whitney U-Test was used to compare the writing performance of the students after the instruction. Table 3 shows the results of the test.

<table>
<thead>
<tr>
<th>Writing Tasks</th>
<th>Groups</th>
<th>N</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
<th>U</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideas/Content</td>
<td>Experimental</td>
<td>12</td>
<td>8.58</td>
<td>103.00</td>
<td>25.00</td>
<td>.03**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>9</td>
<td>14.22</td>
<td>128.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>Experimental</td>
<td>12</td>
<td>8.79</td>
<td>105.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>9</td>
<td>13.94</td>
<td>125.50</td>
<td>27.50</td>
<td>.05*</td>
</tr>
<tr>
<td>Organization</td>
<td>Experimental</td>
<td>12</td>
<td>9.50</td>
<td>114.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>9</td>
<td>13.00</td>
<td>117.00</td>
<td>36.00</td>
<td>.19</td>
</tr>
<tr>
<td>Writing Rules</td>
<td>Experimental</td>
<td>12</td>
<td>9.00</td>
<td>108.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>9</td>
<td>13.67</td>
<td>123.00</td>
<td>30.00</td>
<td>.08*</td>
</tr>
<tr>
<td>Writing</td>
<td>Experimental</td>
<td>12</td>
<td>8.25</td>
<td>99.00</td>
<td>21.00</td>
<td>.01**</td>
</tr>
<tr>
<td>Performance</td>
<td>Control</td>
<td>9</td>
<td>14.67</td>
<td>132.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As is seen in Table 3, the Mann-Whitney U values (U = 21.00, p < .05) for the writing performance of the experimental and control groups indicate that there is a statistically significant difference between the scores of the groups in favor of the control group. According to the results related to the sub-skills of writing performance, it is seen that students in the control group have statistically significantly higher scores in the aspect of ideas/content (U = 25.00, p < .05). In contrast, no statistically significant difference was detected between the scores in the aspects of fluency (U = 27.50, p > .05), writing rules (U = 36.00, p > .05), and organization (U = 30.00, p > .05). It can be inferred from the results that writing instructions in an individual weblog-integrated environment based on blended instructional design is more effective than writing instructions in a cooperative weblog-integrated learning environment based on blended instructional design, in terms of students’ writing performance. The anticipated reasons for this result are discussed in the “Results” chapter.
Results of the qualitative aspect

The following two questions were asked to students on the interview.

- Would you prefer cooperative or individual weblog writing activity if you had the opportunity to choose?
- What are your opinions about the effects of group work self-assessment, group work peer assessment, and teacher assessment concerning the effect of giving feedback on weblog on writing performance?

Results presented under the two themes which is following these questions and supported with concept maps.

The effect of cooperative weblog writing on writing performance

Twelve students were asked first question. Five students gave no response. As seen in Figure 1, three students preferred cooperative weblog writing and four students preferred individual weblog writing. The fact that only three students preferred cooperative weblog writing supports the results of the quantitative aspect.

Three students who preferred cooperative weblog writing stated that cooperative learning was more effective on their writing performance and was more personally useful. They also expressed that they had a chance to exchange information and to acquire more detailed information; in addition, while exchanging information, a discussion platform was created in which, as students stated, learning was facilitated. The examples below reflect the views of the students who preferred cooperative writing.

Student 5: …we had more information in the groups in which we discussed the same topics. We learned more detailed information from our friends.

Student 1: Working with a group was like learning by discussion.

Students who stated that if they had been given an opportunity to choose, they would have chosen individual weblog writing instead of cooperative writing, mentioned the disadvantages of cooperative working. They said that during group discussions, they encountered conflicts and some of the group members showed an intolerant attitude towards each other by dominating others. The following example reflects the views of students who preferred individual writing.
Student 4: Writing with a group was really difficult because everybody in the group wanted something different. Everybody wanted to do keyboard typing.

Student 9: For example, in our group someone wanted us to write what s/he said all the time and didn’t accept our ideas. It was really hard and annoying.

When the students were asked to mention what they thought about the impact of teacher’s guidance, nine students did not give any response, while, as seen in Figure 2, three students mentioned the positive impact of the teacher’s guidance.

According to the students, teacher guidance had a positive effect on conflicts, but teachers could be more helpful as the number of groups decreases. In addition, one of the students indicated that the problems encountered during the performance process could have been solved without teacher guidance by simply thinking. The example below reflects the students’ views on teacher guidance.

Student 12: When we had problems, we had a chance to share them with our teacher.

Student 1: Actually, when there was a problem we could solve it by thinking/solve it on our own.

The effect of giving feedback on a weblog on writing performance

Students were asked second question. When the students were asked about the effect of group work self-assessment, as can be seen in Figure 2, two students expressed positive views.

![Diagram](image)

**Figure 2.** Students’ views on the effect of feedback on writing performance

Students believe that group work self-assessment helps them notice their errors and it improves their writing performance when it is done regularly. Students’ views on the group work self-assessment are as follows:

Student 2: I realized that we wrote better as we did more assessments.

Student 11: While completing the form, we noticed the mistakes we made as we were writing the previous story. We tried to write better.
In order to reveal students’ views on the effect of group work peer assessment on writing performance, students were asked about the effect of the comments that their friends wrote on their story and the effect of their own comments on their friends’ stories. As can be seen in Figure 2, three students expressed positive views and two students expressed negative views on the effect of peer assessment.

In the context of the effect of group work peer assessment, students were asked about the effect of the comments written by their friends on their writing performance. One student expressed a positive view and two students expressed negative views. The student who made a positive comment said that the comments of their friends from different groups helped them notice their errors. The following example reflects the positive view of a student:

Student 12: The good thing about their comments is that they highlight all the errors we’ve made.

Despite this view, two students stated that their friends’ comments on the weblog created a negative effect on their writing performance. The students found the comments insufficient as they were not detailed enough and that they repeated themselves. The comments of these students are as follows:

Student 5: I didn’t like the comments that our friends wrote for us. I expected them to correct our writing errors. Their comments were too short. They should have shown our errors in a better way.

Student 12: Some of our friends wrote the same comments all the time.

When the students were asked about the effect of their comments on weblog on their friends’ writing performance, two students expressed positive views. None of the students expressed negative views. The students noted that they were able to share their experiences, in addition to helping their friends make fewer errors with the help of comments written on the stories. The following example reflects the positive view of one of the students:

Student 9: Our friends started to pay more attention to their errors. We showed them how to write. It helped them.

The students were also asked about the effect of the teacher’s comments on their writing performance and motivation. As can be seen in Figure 2, three students stated that the teacher’s comments helped them notice the points that require more attention and helped them write more elaborately. In addition, according to them, their teacher’s comments were more serious than their friend’s comments. The students’ views are as follows:

Student 3: I care about teacher’s comments more. I read my friends’ comments, too and kept them in mind while writing our next story. I warned my friends. The teacher writes better comments, though.

Student 5: Well, after the teacher’s comments, we paid more attention to writing rules like writing -de, -ki (commonly confused suffixes in Turkish). It was helpful.

**Discussion and conclusions**

According to the results of the quantitative aspect of the research, writing instruction in an individual weblog-integrated learning environment based on blended instructional design, in terms of the ideas/content aspect, contributed to the improvement of writing performance more than writing instruction in a cooperative weblog-integrated learning environment.

The reason for this result is that an individual weblog-integrated writing instruction environment based on blended instructional design could present a blog page in which students could freely make designs. This opportunity may have motivated the students to create more original ideas in their stories and to be more ambitious in content writing. The results that writing with blogs improves thinking skills that were reached by different researchers (Franklin-Matkowski, 2007; Miller, 2011; Olander, 2007) who study the effect of blogs on writing performance and motivation parallels this idea. Additionally, Du and Wagner (2007) mention that individual weblogs have the nature that reinforces taking responsibility and the opportunity to compare the weblog writings. Similarly, Jones (2006) supports this idea with the result of her study, which focuses on the advantages of being the owner of the content that she has created through individual writing studies. Thus, this opportunity which was presented by “individual weblog
environment” could provide much more meaningful improvement for writing performance comparing to “cooperative weblog environment.” Olthouse (2010), who studied the effect of different contexts of writing performance, supports the idea with the result about writing performance was effected from context.

Other reason for this result is -as students mentioned in the focus group interview- the ideas of some of the students who worked in cooperative groups in a weblog environment were not accepted by the others. Because of this, students in the “cooperative weblog environment” might not have expressed their ideas independently in comparison to the students in the “individual weblog environment.” This limitation may have also prevented the development of skills to link the ideas within and between paragraphs. This finding supports the positive effects of an “individual weblog environment” on developing ideas/content performance, contrary to a “cooperative weblog environment.” Additionally, the conflicts among the group members explain the fact that a “cooperative weblog environment” has no effect on ideas/content performance. This was also prevented students from focusing on the aims of the study and to express themselves clearly. During the focus group interview, most of the students stated that they would have preferred doing the weblog writing activities individually.

According to other results revealed in the quantitative process, writing instruction in a cooperative weblog-integrated learning environment based blended instructional design is not more efficient than writing instruction in an individual weblog-integrated learning environment in terms of improving the performance on fluency, organization and writing rules tasks. The first explanation for this may be the conflict problems which were expressed in the focus group interview. For instance, the fact that students could not reach a consensus while planning their writing and did not effectively use the time given by their teacher may have made it difficult to improve their abovementioned skills. Additionally, it was revealed that the egocentric and intolerant attitudes of some of the group members negatively affected students’ performance. So, the conflicting views of students on the effect of cooperative planning and cooperative weblog writing on writing performance support the finding that cooperative weblog writing is not effective for performance improvement on the three abovementioned tasks.

The second explanation for the results could be attributed to the fact that there is no significant effect of an “individual weblog environment” on the improvement of performance on three above mentioned tasks may be the students’ age. In other words, 5th grade students might not have taken enough responsibility for cooperative learning in the weblog environment because of their age. According to the results of the qualitative aspect of the study, students found the comments on their stories insufficient. The fact that the related researches (Sun, 2010; Franklin-Matkowski, 2007; Miller, 2011; Blankenship, 2007) that examined the development of writing performance on the weblog were conducted mainly on high school and university students supports this argument.

In addition, the fact that both of the environments do not have a significant effect on improving the performance related to the three abovementioned writing tasks is similar to the results of Clark’s (2009) research in which weblog and social network writing studies were conducted on students aged 12-13 years. Clark stated that youngsters cannot make compound sentences, do not have a wide range of vocabulary or use a variety of words in their writings. Furthermore, they do not pay attention to capitalization, punctuation, and spelling.

Finally, according to the results of the quantitative aspect of the research, writing instruction in an individual weblog-integrated learning environment based on blended instructional design is, as a whole, more effective on writing performance compared to the writing instruction in a cooperative weblog-integrated learning environment. This can be attributed to the fact that individual learning is more appropriate for the characteristics of writing studies than cooperative learning.

When weblog writing studies are examined (Blankenship, 2007; Chen et al., 2011; Miller, 2011; Sun, 2010), it can be seen that the studies were conducted individually. In the studies mentioned above, it was revealed that an “individual weblog environment” is effective on writing performance. This finding is parallel to the findings of this research in terms of improving the performance on ideas/content tasks and writing performance as a whole.

Due to the fact that the quantitative aspect of this research is experimental and it is not possible to generalize the results, as the research has a qualitative aspect, in order to improve writing performance, which has lifelong importance, the effects of similar studies concerning the first stage of primary school education should be examined.
In the light of the limitations of this study, suggestions for the new research are as follows: More time should be allocated to plan cooperative weblog-integrated writing instruction that is based on blended instructional design. In addition, to increase academic achievement and performance, the effects of a weblog environment should be examined in the process of the curriculum development of all subjects.

Acknowledgments

We would like to thank to Ali Baykal and Serap Emir, who are the thesis jury members, and have contributed to the development of the dissertation which this study is derived from, for their valuable opinions.

References


Santangelo, T., & Olinghouse N. G. (2009). Effective writing instruction for students who have writing difficulties. Focus on Exceptional Children, 42(4), 1-20.


The Role of Group Interaction in Collective Efficacy and CSCL Performance

Shu-Ling Wang1*, Hsien-Yuan Hsu2, Sunny S. J. Lin3 and Gwo-Jen Hwang1

1Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // 2Department of Leadership and Counselor Education, University of Mississippi, Mississippi, USA // 3Institute of Education, National Chiao-Tung University, Taiwan // shuling@mail.ntust.edu.tw // suhy0914@gmail.com // sunnylin@mail.nctu.edu.tw // gjhwang.academic@gmail.com

*Corresponding author

(Submitted June 8, 2013; Revised December 25, 2013; Accepted March 14, 2014)

ABSTRACT

Although research has identified the importance of interaction behaviors in computer-supported collaborative learning (CSCL), very few attempts have been made to carry out in-depth analysis of interaction behaviors. This study thus applies both qualitative (e.g., content analyses, interviews) and quantitative methods in an attempt to investigate the role of interaction behaviors (i.e., cooperative processes, cognitive involvement) in group motivation (i.e., collective efficacy) and group performance in CSCL. A total of 35 college students generated 387 cooperative processes and 421 cognitive ideas during the collaborative process over six weeks. The results indicated that the quality of interaction behaviors was positively related to CSCL performance. High performance groups were more involved in complex cooperative processes, and expressed more cognitive ideas at both high and low cognitive levels than low performance groups. For the role of interaction in collective efficacy, the quantity rather than quality of interaction behaviors played a more critical role in constructing collective efficacy. High efficacy groups had more cooperative processes and cognitive ideas than low collective efficacy groups. The interviews also revealed that group interaction played a mediating role in constructing collective efficacy. For the role of collective efficacy in group performance, the results also show that the high efficacy groups performed better than the low efficacy groups. Implications and suggestions for future research are also provided.

Keywords

Group interaction, Collective efficacy, Computer-supported collaborative learning (CSCL), Cooperative process, Cognitive involvement

Introduction

The development of technology has been found to have positive effects on student learning. Particularly, research suggests that collaborative learning with computer technology (e.g., CMC) or computer-supported collaborative learning (CSCL) has gained great attention in the educational context, because it provides students with more flexible ways of acquiring information (Palmieri, 1997; Wang & Lin, 2007). In a meta-analysis of 122 studies with technology in CSCL (hereafter, CSCL is used for all terms regarding collaborative learning with technology), the results in general support that collaborative learning promotes students’ use of strategies, positive attitudes and achievement (Lou, Abrami, & d’ Apollonia, 2001). However, research also suggests that students may behave differently in their interaction, such as in online discourse (Caspi, Chajut, Saporta, & Beyth-Marom, 2006; De Laat & Lally, 2003; Häkkinen & Järvelä, 2006; Liu, Chung, Chen, & Liu, 2009; Salovaara & Järvelä, 2003), and cognitive activities (De Laat & Lally, 2003; Hurme, Palonen, & Järvelä, 2006; Ke, 2013). For example, research has proposed students’ behavioral problems in CSCL discussion, including difficulties in connecting online comments to course concepts, providing assertions without evidence, being less critical online (Murray, 2000), and generating less high quality discourse regardless of education level (Lipponen, Rahikainen, Hakkarainen, & Palonen, 2002; Wang & Hwang, 2012). To better understand students’ interaction behaviors, this study thus attempts to apply qualitative analyses to students’ authentic interaction behaviors occurring in CSCL, which should provide a more comprehensive understanding of students’ interaction behaviors, and along their role in CSCL performance.

In addition to group interaction, research has indicated that group motivation, namely, collective efficacy, has positive effects on group performance in various areas, including schools, organizations, and sports (Bandura, 1997; Goddard, 2001; Klassen & Krawchuk, 2009; Schunk, Pintrich, & Meece, 2008; Wang & Lin, 2007). However, very few attempts have been made to examine the role of collective efficacy in CSCL performance. This study attempts to fill this gap. In addition, research has shown that collective efficacy is critical in collaborative learning, so there is a need to learn more about its construction. Bandura (1997) proposes the theory of collective efficacy and also suggests that group interaction is very likely to influence the construction of collective efficacy. As research...
suggested, CSCL environments containing less social clues, such as body gestures, and facial expressions, result in various interaction behaviors as compared to traditional environments (Wang & Hwang, 2012). Group interaction may play a particular role in collective efficacy in the CSCL environment. Hence, this study also attempts to investigate the role of interaction behaviors in constructing collective efficacy in the CSCL environment.

**Theoretical background**

*Group interaction (cooperative processes, cognitive involvement), group performance and collective efficacy*

Recent research has shown that interaction plays a significant role in CSCL (Hooper, 2003; Ke, 2013; Jung, Choi, Lim, & Leem, 2002; Puntambekar, 2006; Van Drie, Van Boxtel, Jaspers, & Kanselaar, 2005), possibly due to the text-based features and flexibility of the online environment, which makes interaction more convenient and conducive to deep thought, but may also be less focused or even uncontrollable. For example, although research suggests that online written discussion requires more focused reflection, and thus may develop deeper thought (Serçe, et al., 2011; Wiley & Schoofer, 2001), other research suggests that students tend to be passive, less effective, and less critical in online discussion (Chiu & Hsiao, 2010; Murray, 2000; Walther, Loh, & Granka, 2005), and produce limited cognitive quality during group interaction in CSCL (Wang & Hwang, 2012). Since the effect of group interaction during discussion in CSCL is not clear, the role of interaction behaviors in CSCL is still in need of further analysis. This study therefore attempts to investigate the quality of interaction behaviors, such as cooperative processes and cognitive involvement, along with their roles in CSCL performance.

With respect to cooperative processes, according to Hertz-Lazarowitz (1995), the quality of cooperative processes depends on the actual interaction behavior that occurs in the three stages of learning: input, process and outcomes. As Hertz-Lazarowitz (1995) suggested, simple cooperative processes occur in the input stage such as giving resources, or in the outcome stage such as providing answers without any further discussion, which are considered as low quality interaction. On the other hand, high quality cooperative processes, such as complex cooperative processes, take place in the process stage, for example, engaging in the process of producing a product. Recent research has indicated that students in online text-based discussion produce relatively more initiatives but fewer responses (Strømsø, Grøttum, & Lycke, 2007), which is in accordance with Hertz-Lazarowitz’s simple and complex cooperative processes, respectively. This study thus attempts to further explore the role of cooperative processes, such as simple and complex processes, in the CSCL environment.

In addition, research has suggested that cognitive involvement is important for learning in computer-mediated communication (Henri, 1992). A recent study indicates that online text construction could be enhanced by students’ metacognition (Yeh & Yang, 2011). Moreover, research also indicates that students who use higher level cognitive strategies have better performance (Garavalia & Gredler, 2002; Wang & Wu, 2008). Willoughby, Wood, McDermott and McLaren (2000) also suggest that, when a group is able to share sophisticated strategic information, the group members are more likely to make contributions to promoting knowledge. This study therefore also strives to explore the quality of cognitive involvement (defined as cognitive ideas applied in learning) and its role in CSCL performance.

Moreover, research indicates that interaction increases student motivation (Jung et al., 2002; Markett, Arnedillo Sánchez, Weber, & Tangney, 2006) and positive interaction promotes group motivation (Johnson & Johnson, 1998). In particular, research suggests that group interaction is more likely to affect collective efficacy (Bandura, 1997, 2000). Jung and Sosik (2003) further illustrate that group members with more frequent interaction tend to develop stronger collective efficacy in the classroom context. Therefore, in addition to examining the role of cooperative processes and cognitive involvement in group performance, this study also investigates the role of these interaction behaviors in collective efficacy in the CSCL environment.

*Collective efficacy and performance*

According to Bandura (1997), collective efficacy is defined as a group’s shared beliefs in its conjoined capabilities to perform the sequences of action required to achieve designated goals. In other words, collective efficacy is concerned with the performance capability of a group as a whole. Research shows that collective efficacy has a significant effect on group functioning, especially on levels of effort, persistence and achievement (Bandura, 1997, 2000;
Research also indicates that collective efficacy is positively correlated to group performance in schools, organizations, and sports (Bandura, 1997; Chow, 2009; Goddard, 2001; Hodges & Carron, 1992). Two meta-analysis studies indicate that the relationship between collective efficacy and group performance is significantly positive (Gully, Incalcaterra, Joshi, & Beaubien, 2002; Stajkovic et al., 2009). For example, a meta-analysis investigating 6,128 groups with 96 studies shows that collective efficacy is significantly related to group performance ($r = .35$) (Stajkovic et al., 2009). This study thus hypothesizes that collective efficacy should have similar effects on CSCL performance.

Research questions

- What is the role of interaction behaviors (i.e., cooperative process, cognitive involvement) in CSCL performance?
- What is the role of interaction behaviors (i.e., cooperative process, cognitive involvement) in collective efficacy in the CSCL environment?
- What is the role of collective efficacy in CSCL performance?

Methodology

Measures

*Questionnaire*

The scale of “collective efficacy” consisted of 8 items (e.g., I believe that my group can understand the most difficult part of this course) and has proven to be very reliable and valid in previous studies (Wang & Lin, 2007). The questionnaire used a 7-point Likert scale. The results of the reliability test indicated that it was very reliable, with an alpha of .926. All items were loaded with a factor loading over .40.

*The networked system*

![Figure 1. The NetPorts system (Wang & Lin, 2007)](image)

This study used the networked portfolio system (NetPorts, Figure 1), which has been proven to be reliable and applicable (Wang & Lin, 2007). The system was designed with a system management module which allows teachers to easily monitor student progress. In addition, the NetPorts system provided an online questionnaire module to collect students’ data, and also a non-synchronous online chat room that allowed students to discuss their group assignment with their group members. Moreover, students were able to upload homework, process peer assessments, propose ideas and view records of self-reflection, peer feedback and ideas using NetPorts.
Participants

There were 35 students generating 387 interaction processes and 421 cognitive ideas over six weeks in this study. The selected participants were students enrolled in a mandatory course entitled “Educational Measurement and Evaluation” in the Teacher Education Center of a research university in Taiwan. The Teacher Education Center offers courses for students from various majors who consider the teaching profession as their future career. Of these 35 participants, 10 were from Mathematics, 15 from Foreign Languages, 3 from Science, 3 from Biology, 2 from Civil Engineering, and 2 from Computer Science.

Procedures

Group assignment

The group assignment was to design test questions for a high school mid-term examination based on the group members’ academic domain. For example, students who were majoring in Biology would write the test items for the mid-term Biology exam. All groups needed to discuss and decide on the content of units and question format (e.g., multiple choice, short answer, true-false questions, etc.), as well as design the test items of the exam in accordance with Bloom's two-way specification table, which also placed demands on their cognitive effort. The design of the mid-term exam was not a simple but an ill-structured task, and should encourage more group interaction and cognitive involvement during the process. Finally, the group assignments were evaluated and graded by the instructor based on the scoring rubrics. Thus, every group would receive a score based on their performance of the assignment (i.e., group performance).

Grouping

Research has suggested that small group size encourages group interaction (Abrami et al., 1996), so each group only consisted of 2-3 students from the same major, since they needed to complete a group assignment which was based on their academic domain. If only two students were from the same major, there were only two students in the group.

Process

The students were requested to discuss their group projects using the NetPorts system over a period of six weeks. The students could discuss their projects in the non-synchronous online chat room through Netports. Students could
only discuss their projects with their group members. They could initiate or reply to a topic issue posted on their group discussion board through NetPorts (see Figure 2). Students’ networked discussion would be considered as their participation score (15%) for their final grade, a form of participation evaluation common in instructors’ grading systems. All students had to fill out the questionnaire of “collective efficacy” through the NetPorts system, and the system would then allow them to upload their group projects.

After completing the collaborative task, researchers also interviewed students from two groups with high quality interactions, and two groups with low quality interactions. These students were interviewed individually. The interview questions consisted of students’ perceptions of group interaction and their influence on collective efficacy, as well as the factors influencing group interaction.

This study used both quantitative and qualitative methods for data analysis.

**Quantitative method**

Statistical techniques were used to analyze the data. Traditionally, a t-test is used to assess the statistical significance of differences between groups. However, according to Coe (2002), “statistical significance” in statistics, the likelihood of the difference between two groups, could just be an accident of sampling, and other researcher has also suggested that Cohen’s effect size, which is independent of sample size, can provide practical significance for better data interpretation (Cohen, 1988). Although there were 387 interaction processes and 421 cognitive ideas for data analysis, there were only 13 groups in this study, for which it was difficult to achieve traditional statistical significance among groups. Thus, the Cohen’s effect size, which is not affected by the sample size, should be more appropriate for the data analysis of this study. In addition, the recent publication manual of the APA (American Psychological Association) (2009) has claimed that “it is almost always necessary to include some measure of effect size in the results section…such as a Cohen’s d value” (p. 34). Thus, in addition to the statistical significance test, this study also reported the practical significance using the Cohen’s d test for data analysis. According to Cohen (1988), the effect size $d = 0.2$ is considered as a small effect size, $d = 0.5$ as medium, and $d = 0.8$ as large. It is important to note that, in this study, only the large effect size ($d \geq .8$) is considered as having practical significance in our data analysis, while small and medium are not.

In addition, although our groups consisted of either two or three members (according to their academic domain), which might be suspected as having an effect on group interaction, after checking all the quantitative results, our results showed no difference between the groups with two members and those with three members, as well as no difference between individual (i.e., group interaction data/members) and group level analysis. Since this study places emphasis on computer-supported collaborative learning, the results are presented at the group level of analysis.

**Qualitative methods**

In addition to the quantitative analysis for questionnaire and hypothesis, this study also applied qualitative methods, such as content analysis of online discussion and interviews to explore more in-depth information. For example, researchers have suggested that content analysis involving both numeric and interpretive data analysis provides a more meaningful analysis of group discussion (Hara, Bonk, & Angeli, 2000; Henri, 1992; Woo & Reeves, 2007), as well as that interviews help gain insight into participants’ experiences (Turner, 2010). In this study, two content analyses were used to analyze the CSCL discourse, one for the students’ cooperative processes, and the other for the quality of cognitive involvement. The analysis of the cooperative processes was based on Hertz-Lazarowitz’s (1995) three phases of interaction (input, process and outcomes). Students’ interaction behaviors (i.e., posted messages) only involved in the input phase (e.g., giving resources), or in the outcome phase (e.g., providing answers) were considered as simple cooperative processes, while interactions engaged in the process of achieving a cooperative goal, such as engaging in discussions, were perceived as complex cooperative processes. Each posted message to other group members only counted as one cooperative process (i.e., a simple or complex process). Therefore, the sum of the simple and complex processes was the total number of cooperative processes of an individual. The total cooperative processes of each group were the total cooperative processes of all group members.
In addition to cooperative processes, the quality of the students’ cognitive involvement (i.e., the quality of cognitive ideas) was also analyzed. Each idea was considered as one unit to be analyzed. The assessment of students’ cognitive ideas during the CSCL discussion was based on Bloom’s (1956) taxonomy of educational objectives. The taxonomy consisted of six cognitive levels, ranging from knowledge, comprehension, application, analysis, synthesis to evaluation. Research suggests that knowledge and comprehension are viewed as low-level cognitive skills, whereas application, analysis, synthesis and evaluation are considered as high-level cognitive skills (Swart, 2010). Thus, student group discussion ideas categorized as knowledge and comprehension were coded as low-level cognitive ideas, whereas ideas comprising application, analysis, synthesis and evaluation were coded as high-level cognitive ideas. It is important to note that our study placed more focus on the role of task-related interactions in CSCL, so only task-related (or cognitive-related) interactions were evaluated for cognitive involvement and cooperative process, while off-task discussions (i.e., discussions not related to task content or cognitive-unrelated messages) were not analyzed in our content analysis. Moreover, the interviews were also transcribed and coded using content analysis.

Two graduate students in the Department of Education served as the raters for this study. The inter-rater reliability was calculated by the Cohen’s Kappa. The results indicated that the inter-rater reliability was good with .826 for the simple processes, .873 for the complex processes, .887 for low cognitive involvement, and .868 for high cognitive involvement. This demonstrated that content analysis was a reliable measure for interaction behaviors in our study.

Results

The descriptive statistics of content analysis for cooperative process and cognitive involvement were shown in Table 1. Taking the average of the scores from the two raters, the descriptive statistics showed that there were a total of 387 cooperative processes ($M = 29.77, SD = 11.80$) and 421.5 cognitive ideas ($M = 32.42, SD = 16.59$) generated by the 13 groups. Our analysis also showed that all variables analyzed in our study were approximately normally distributed.

<table>
<thead>
<tr>
<th>Cooperative processes</th>
<th>$\bar{X}$</th>
<th>$SD$</th>
<th>Max</th>
<th>Min</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex processes</td>
<td>29.77</td>
<td>11.80</td>
<td>48.00</td>
<td>10.00</td>
<td>387</td>
</tr>
<tr>
<td>Cognitive ideas</td>
<td>4.77</td>
<td>5.08</td>
<td>16.50</td>
<td>0.00</td>
<td>62</td>
</tr>
<tr>
<td>High cognitive ideas</td>
<td>32.42</td>
<td>16.59</td>
<td>59.00</td>
<td>11.00</td>
<td>421.5</td>
</tr>
<tr>
<td>Low cognitive ideas</td>
<td>6.65</td>
<td>5.69</td>
<td>16.50</td>
<td>0.00</td>
<td>86.5</td>
</tr>
</tbody>
</table>

Examples of content analysis for the cooperative processes were shown in Table 2. As previously stated, a simple cooperative process refers to a group member’s posted messages in either the input or the output phase, without any on-task discussion, while a complex cooperative process refers to group members’ messages engaged in the on-task discussion, usually with a discussion thread. For example, a group member replying to others with, “I fully agree with your viewpoint, I think we should make the test questions in accordance with every educational objective…” is considered as a complex cooperative process. The pair t-tests showed a significant difference ($t = 8.16, p < .01$) between the simple ($M = 25, SD = 9.15$) and complex cooperative processes ($M = 4.77, SD = 5.08$); that is, students had significantly more simple cooperative processes than complex ones.

<table>
<thead>
<tr>
<th>Examples of cooperative process</th>
<th>Mean</th>
<th>$SD$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple process</td>
<td><strong>25.00</strong></td>
<td>9.15</td>
<td>8.16</td>
</tr>
<tr>
<td>I found the following test item examples on the Internet, please refer to the attached files (only input phase, no discussion).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have completed the short-answer questions for our projects, so I have done my work, please just add it in (output phase, no discussion).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex process</td>
<td>4.77</td>
<td>5.08</td>
<td></td>
</tr>
<tr>
<td>I fully agree with your viewpoint, I think we should make the test questions in accordance with every educational objective,…so I will check every test question and its corresponding educational objective. (process phase, with discussion thread)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .01*
In addition, for the content analysis of cognitive involvement, Table 3 showed both students’ low and high cognitive ideas in their group discussion. For example, one student posted the message “For the items you constructed, the most problematic issue is that you used a multiple choice format to assess students’ cognition of evaluation level. Multiple-choice could be adequately used to measure lower level cognition but not higher level cognition...”, which was considered as analysis level of cognition, and coded as a high cognitive idea. The pair t-tests showed a significant difference ($t = -6.83, p < .01$) between low ($M = 25.77, SD = 12.49$) and high cognitive ideas ($M = 6.65, SD = 5.69$); that is, students generated significantly more low cognitive ideas than high cognitive ideas.

### Table 3. T-test and examples of content analysis for cognitive ideas

<table>
<thead>
<tr>
<th>Cognitive Levels</th>
<th>Examples of cognitive levels in content analysis</th>
<th>M</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>According to the textbook, making short-answer questions for the test is the easiest.</td>
<td>25.77</td>
<td>12.49</td>
<td>-6.83**</td>
</tr>
<tr>
<td>Comprehension</td>
<td>I think making a good test should consider 4 sources that influence quality: reliability, validity, referencing, and objectivity. Among these, reliability and validity are of higher priority for quality test construction.</td>
<td>25.77</td>
<td>12.49</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>In constructing a multiple choice item, the plausibility (seeming likely to be true for test takers) of each choice is my priority concern. Thus, the items that I constructed are listed in the following....</td>
<td>6.65</td>
<td>5.69</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>For the items you constructed, the most problematic issue is that you used a multiple choice format to assess students’ cognition of evaluation level (in Bloom’s taxonomy). Multiple-choice could be adequately used to measure lower level cognition, but not higher level cognition. Therefore, I suggest that you construct some “subjective scored” items such as essays.</td>
<td>6.65</td>
<td>5.69</td>
<td></td>
</tr>
<tr>
<td>Synthesis</td>
<td>Short-answer questions cannot be used for complex or high-level learning outcomes, so I suggest integrating the short-answer questions into the authentic assessment.</td>
<td>6.65</td>
<td>5.69</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>Your designed test seems not to be in accordance with the rule of the “two-way specification table”. Based on the rule, the item construction should consider both the educational objective and content, so the ratio of item constructions should be .....</td>
<td>6.65</td>
<td>5.69</td>
<td></td>
</tr>
</tbody>
</table>

** $p < .01$

The results for the study’s research questions are showed as follows:

**What is the role of interaction behaviors (i.e., cooperative process, cognitive involvement) in CSCL performance?**

To better understand the role of interaction behaviors in CSCL performance, we divided the groups into high and low performance clusters based on the scores of group assignment. The scores of 13 groups ranged from 63.00 to 87.00 with mean of 74.77 and SD of 7.64. The high performance cluster comprised groups of performance scores in the
approximate top 40%, while the low performance cluster consisted of groups of scores in the bottom 40%. For both high and low performance clusters, our analyses showed that simple/complex processes and low/high cognitive ideas were nearly normally distributed. The t-test results indicated a significant difference between the high and low performance clusters ($p < .01$). For the role of cooperative process in group performance, as shown in Table 4, both the t-test ($t = -2.36$, $p < .05$) and Cohen’s $d$ of 1.50 showed that the higher performance groups engaged in more complex cooperative processes than the low performance groups, but there was no difference in the simple cooperative processes between these two groups. That is, the high performance groups had more complex cooperative processes than the low performance groups.

For the role of cognitive involvement in performance, although the t-test results only showed a significant difference for low cognitive ideas, the Cohen’s $d$ values showed practical differences for both low (Cohen’s $d = 2.29$) and high level (Cohen’s $d = 1.01$) cognitive ideas between the low and high performance groups. Our results also indicated that the correlation between high cognitive ideas ($r = .62$) and performance, as well as low cognitive ideas ($r = .84$) and performance were both significant at the .05 level. In other words, both high and low cognitive ideas were significantly related to CSCL performance. It seems reasonable that students not only need to apply high level cognitive ideas but also low level ideas to achieve a better outcome, because a low cognitive level of information processing, such as knowledge and comprehension, is the basis of knowledge construction. To better understand students’ total cognitive involvement (the sum of the low and high level cognitive ideas) in learning, our t-test ($t = -3.38$, $p < .01$) and Cohen’s $d$ ($d = 2.13$) results also indicated a significant difference in the total cognitive involvement between the low and high performance groups. In other words, the high performance groups had more cognitive involvement in group discussion than the low performance groups.

<table>
<thead>
<tr>
<th>Table 4. The comparison of interactions between high and low achieving groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Simple processes</td>
</tr>
<tr>
<td>Complex processes</td>
</tr>
<tr>
<td>Total cooperative processes</td>
</tr>
<tr>
<td>Low cognitive ideas</td>
</tr>
<tr>
<td>High cognitive ideas</td>
</tr>
<tr>
<td>Total cognitive ideas</td>
</tr>
</tbody>
</table>

$t$ value: $p < .05$. $^{*}p < .01$; Cohen’s $d$: $^*$large effect size

**What is the role of interaction behaviors (i.e., cooperative process, cognitive involvement) in collective efficacy in the CSCL environment?**

To better understand the role of interaction behaviors in collective efficacy, students were also divided into high (top 40%) and low (bottom 40%) collective efficacy clusters. The data of both high and low collective clusters appeared approximately normally distributed in our analysis. The result of a Cohen’s $d$ value of 1.36 showed a practical significance between high and low collective efficacy clusters, and the t-test results almost achieved a significant difference between these two clusters at the 0.05 level ($t = -2.14$, $p = .06$). For the cooperative process, as shown in Table 5, the Cohen’s $d$ value of 1.44 indicated a large effect size in the simple cooperative process between low and high efficacy groups, but no difference in the complex cooperative process between these two groups. The results also revealed that there were practical significant differences (Cohen $d = 1.31$) in the total cooperative processes (the sum of the simple and complex cooperative processes). That is, the high collective efficacy groups had more simple and total cooperative processes than the low collective efficacy groups.

For the role of cognitive involvement in collective efficacy, the Cohen’s $d$ values of 1.30 and 1.18 revealed a large effect size for low cognitive ideas as well as total cognitive ideas, and .46 was close to the medium effect size for high cognitive ideas between the low and high efficacy groups.

<table>
<thead>
<tr>
<th>Table 5. The comparison of interactions between high and low collective efficacy groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Simple processes</td>
</tr>
<tr>
<td>Complex processes</td>
</tr>
<tr>
<td>Total cooperative processes</td>
</tr>
</tbody>
</table>

249
The interview analysis also showed the relationship between group interaction and collective efficacy. In the interviews with the two high quality interaction groups, the students reported that their groups had high confidence in achieving their group goal through members’ involvement in the group discussion. On the other hand, the students in the low interactive groups reported that they did not have confidence in completing the group project due to their team members’ low interaction and commitment in the group discussion. Interestingly, in the low interactive groups, the interviews also found that even though the students initially perceived themselves or their team members as having high capability or high efficacy (i.e., self-efficacy) beliefs, without interaction, the group members eventually came to hold low collective efficacy, believing that their groups could not complete the project with good quality. That is, group interaction played an important role in forming collective efficacy.

Additionally, the students of the high quality groups reported that they had a good leader who played an important role in their group interaction, such as assigning the work, monitoring the process, as well as actively promoting group discussion, which thus enhanced their group’s efficacy beliefs. In other words, leadership may play an important role in collective efficacy, but needs to go through group interaction, which in turn leads to increased collective efficacy.

What is the role of collective efficacy in CSCL performance?

For the role of collective efficacy in collaborative performance, the Cohen’s d value of 1.22 showed a large effect size in achievement between the low ($M = 69.20$, $SD = 4.44$) and high collective efficacy groups ($M = 77.60$, $SD = 8.56$). In other words, the Cohen’s d value showed that high collective efficacy groups perform better than low collective efficacy groups.

Conclusions and discussion

In summary, for the role of interaction behaviors in group performance, the results show that the quality of interaction behaviors is positively related to CSCL performance. The higher performance groups engage in more complex cooperative processes, and also appear to express more cognitive ideas at both high and low cognitive levels than the low performance groups. For the role of interaction behaviors in collective efficacy, our results indicate that the quantity instead of the quality of interaction plays a more substantial role in constructing collective efficacy. Our results show that groups with more cooperative processes, and those applying more cognitive ideas tend to have higher collective efficacy. The interview results show a similar finding that collective efficacy is more likely conveyed by group interaction. For the role of collective efficacy in collaborative performance, our results indicate that groups with higher collective efficacy demonstrate better academic performance than low collective efficacy groups in CSCL.

Consistent with our hypothesis, the results show that group interaction is important for students’ CSCL performance. Particularly, our results demonstrate that high performance groups are more involved in group processes rather than only in the input or output stage. Accordingly, a study of the effects of interaction on collaborative concept mapping also demonstrates that complex cooperative processes have positive effects on group performance (Chiu, Huang, & Chang, 2000). Additionally, other researchers also indicate that group members engaging more in group processing develop better understanding (Puntambekar, 2006) and have better academic achievement (Hooper, 2003; Kutnick, Ota, & Berdondini, 2008; Stamovlasis, Dimos, & Tsaparlis, 2006). Indeed, students who are more involved in group processes should gain more feedback from in-depth discussion to improve their group productivity. To promote complex cooperative behavior, Kreijn, Kirschner and Jochems (2003) have suggested that positive interdependence is essential to promote interaction. Teachers may need to provide tasks (e.g., ill-structured tasks) that foster positive interdependence, which require all team members to contribute efforts to achieve their common goal. In addition, the task difficulty should also be taken into account; if the designed task is too simple for the team members, they may complete it individually and provide the final product in the output phase. Moreover, our interviews indicate that
leadership facilitates collaborative process, such as assigning the work, monitoring the process, and promoting group interaction. Therefore, teachers may consider to assign the appropriate leaders to the groups in need in order to promote collaborative process.

In addition, in accordance with other research findings that interaction has influences on motivation (Markett, Arnedillo Sánchez, Weber, & Tangney, 2006; Kutnick, Ota, & Berdondini, 2008), our results further indicate that group interaction is also important in constructing collective efficacy in the CSCL environment. However, our study shows that the quantity rather than the quality of interaction plays a more critical role in forming collective efficacy. As previously mentioned, our results show that both complex processes and high cognitive ideas are far less (4-5 times) than simple processes and low cognitive ideas, respectively. As a result, students’ collective efficacy beliefs are less likely to be affected by very few interactions, such as complex processes or high cognitive ideas in this study. Both these results consistently show that collective efficacy cannot be developed without sufficient interaction. Our study further indicates that groups with more interactions, such as cooperative processes or more cognitive ideas, have higher collective efficacy. However, Chow’s (2009) recent study in sports shows different findings, suggesting that collective efficacy may not be influenced by individuals’ frequent interactions with proximal members or group communication density. This difference is possibly due to the fact that our study examines students’ actual interaction behaviors occurring in CSCL, while Chow’s (2009) study using social network analysis is based on every member’s general perception of their interactions with each other. The investigation of students’ authentic group interactions should help to make our study more legitimate and persuasive in understanding the role of group interaction in the construction of collective efficacy. In addition, it is important to note that the quantity of interaction is particularly important in constructing collective efficacy in the CSCL environment, which in general lacks social presence, gesture, or expression, so students need more interactions to construct their collective efficacy. Teachers may particularly need to encourage more interactions in online environments in order to promote students’ collective efficacy in collaborative learning.

Our results based on content analysis indicates the importance of group interaction in the construction of collective efficacy, and the interview results not only validate these findings but also provide more in-depth information, such as revealing the mediated role of group interaction in constructing collective efficacy. For example, our results indicate that leadership exerts its influence on collective efficacy through group interaction. The results also further reveal that the groups benefit from high efficacy (i.e., self-efficacy) members only when these members are involved in group interaction. In other words, the factors influencing collective efficacy may possibly need to go through group interaction. These aforementioned findings are consistent with Bandura’s (1997) viewpoint; that is, group interaction plays an important role in developing collective efficacy. Future research on the factors that affect collective efficacy should consider the possible mediated effect of group interaction between the factors and collective efficacy.

The limitation of this study is that several factors that may influence group interaction are not taken into account. For example, research suggests that heterogeneous group composition is more conducive to collaborative learning (Webb & Palincsar, 1996), but the group composition of our study was more homogeneous than heterogeneous, because our task was designed for the students to develop an exam based on their academic domain. Although such group composition and task design was more realistic in practice, and might help the students to develop instructional skills for their future career, the homogeneity of the group composition might have some influences on the group interactions. In addition, since our study only focused on the interaction behaviors in an asynchronous environment, the role of synchronous features in group interaction is still unknown. Thus, future studies should take these variables into account, and investigate their influences on group interaction in the CSCL environment. Moreover, a recent study also indicates a strong relationship between collective efficacy and team performance, but further suggests that the relationship is likely moderated by collective efficacy dispersion (i.e., heterogeneous collective efficacy within teams) (Chow, 2009). Thus, the role of collective efficacy believes dispersion may need to be further investigated in the CSCL environment in order to better understand the relationship between collective efficacy and CSCL performance. Furthermore, our interviews also reveal two possible factors influencing group interaction, such as group members with high capability and off-task social-emotional discussion; that is, group members with high capability tend to devote more resources in group interaction, and off-task social-emotional discussions also influence task-related group interaction. These two factors may need to be further examined in future research. Finally, future research should also explore other interaction styles or behaviors (e.g., helping behaviors) in order to better understand their effects on computer-supported collaborative learning and achievement.
Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 99-2511-S-011-009-MY3 and NSC 102-2511-S-011 -007 -MY3.

References


Ke, F. (2013). Online interaction arrangements on quality of online interactions performed by diverse learners across disciplines. Internet and Higher Education, 16, 14-22.


Exploring the Moderating Role of Self-Management of Learning in Mobile English Learning

Rui-Ting Huang
Department of Business Administration, National Chung Hsing University, Taiwan // rthuang0324@dragon.nchu.edu.tw

(Submitted June 30, 2013; Revised December 16, 2013; Accepted January 27, 2014)

ABSTRACT
Although a considerable number of studies have revealed that self-management of learning (SML) could be closely related to learning achievements, there is still a paucity of research investigating the moderating effect of self-management of learning on mobile learning outcomes. Accordingly, the primary purpose of this study was to explore the moderating role of self-management of learning in mobile English Learning. The participants of this study were 389 undergraduate students who used to use handheld electronic dictionaries to learn English before. The Partial Least Squares (PLS) analysis, a component-based Structural Equation Modelling (SEM) technique, was adopted to examine the data in this study. It has been demonstrated that perceived usefulness and playfulness of electronic dictionary could be positively related to mobile English learning satisfaction. Additionally, the study results have revealed that with particular respect to learners with higher SML, resistance to change could have no influence on mobile English learning satisfaction. Finally, it has been found that the self-management of learning could moderate the relationships between key mobile English learning determinants, satisfaction, and continuance intention.

Keywords
Mobile English learning, Self-management of learning

Introduction
Mobile learning, which generally refers to learning activities via the use of mobile devices such as notebook computer, mobile phone, and personal digital assistants (PDA), has gradually played a critical role in helping people acquire new knowledge and skill in life (Valk, Rashid, & Elder, 2010). Probably because the applications of mobile devices are increasing and expanding in today’s teaching and learning environments (Cavus & Ibrahim, 2009; Clough, Jones, McAndrew, & Scanlon, 2008; Goh & Kinshuk, 2006), recently, there has been a growing interest in mobile learning studies (Roca & Gagné, 2008; Wang, Wu & Wang, 2009). Nevertheless, more research is still needed on evaluating mobile learning outcomes (Sung & Mayer, 2013; Uzunboylu, Cavus, & Ercag, 2009). With particular respect to the link between self-management of learning and mobile learning outcome, although a considerable number of studies have revealed that self-management of learning could be closely related to learning achievements (Abar & Loken, 2010; Dignath, Buettner, & Langfeldt, 2008; Kaufman, Zhao, & Yang, 2011), there is still a paucity of research investigating the moderating effect of self-management of learning on mobile learning outcome. More specifically, whether learners with different levels of self-management of learning could have different levels of relationships between key mobile learning determinants, satisfaction and continuance intention has not yet been fully investigated in present studies.

In addition, previous research has indicated that the continuance intention of customers could be a key measure to evaluate the final success of new products or services (Lin, 2012). Although numerous researchers have highly focused on mobile learning studies, relatively little attention has been paid to the role of moderating variables in continuance intention (Lin, 2011). In order to close the gaps in research, and further enhance mobile learning effectiveness and efficiency, it is important that researchers and practitioners in the mobile learning field should concentrate not only on mobile technology adoption, but also on mobile learning outcomes. As mobile technology has gradually and closely connected with our life, the applications of mobile technology in teaching and learning environments will become more common than previously thought. Accordingly, the primary purpose of this study is to explore the moderating role of self-management of learning in mobile English Learning.
Theoretical background and hypothesis development

Mobile English learning and handheld electronic dictionary

Handheld electronic dictionaries have gradually become one of the critical tools for learning English in Chinese speaking countries, probably owing to the convenience of mobile technology and the useful function of electronic dictionaries. Chen (2010) added that “there are no significant differences between pocket electronic dictionary (PED) and paper dictionary (PD) use in comprehension, production and retention of vocabulary although the speed of the former is significantly faster than the latter” (p. 275). More specifically, the potential advantages of using handheld electronic dictionaries include expansive vocabularies, synonyms and grammar references, and as compared to book-form dictionaries, digital-form electronic dictionaries not only allow learners to easily and quickly search for specific words, but also provide them with powerful pronunciation functions that help learners to improve their English ability (Chen, 2010). In view of the critical influence of mobile technology on the quality of language learning, it has been shown that there is a growing interest in mobile assisted language learning studies (Chen & Chung, 2008; Lu, 2008). However, limited studies have been done to examine the moderating role of self-management of learning in mobile English learning satisfaction and continuance intention.

In previous research, it has been well documented that there is a positive correlation between consumer satisfaction and their continuance intention to adopt new IT products or services (Lin, 2012; Zhao & Lu, 2012). That is, as customers and users are more satisfied with new products or services, it is likely that they will have more positive continuance intention. In the same vein, it is conceivable that learners with higher levels of mobile English learning satisfaction (MELS) will have more positive mobile English learning continuance intention (MELCI), which refers to their continued intention to take mobile learning. Consequently, this study proposes the following hypothesis.

H1: Mobile English learning satisfaction (MELS) could have a positive influence on mobile English learning continuance intention (MELCI).

Perceived usefulness (PU)

The perceived usefulness (PU), which refers to “the degree to which a person believes that using a particular system would enhance his or her job performance” (Davis, 1989, p. 320), could play a key part in determining user satisfaction (Roca, Chiu, & Martínez, 2006). That is, in terms of using new information technology, if customers perceive higher usefulness of information technology, it is probable that they will be more satisfied with new IT products or services. Previous online learning studies have shown that the perceived usefulness, which is viewed as extrinsic motivation (Davis, Bagozzi, & Warshaw, 1992), could have a positive impact on learner satisfaction (Arbaugh, 2000; Roca et al., 2006; Sun, Tsai, Finger, Chen, & Yeh, 2008). Nonetheless, a recent report by Kim (2010) has indicated that consumers’ perceived usefulness of mobile data service could not be related to their satisfaction. In mobile learning environments, it is assumed that learners with better perceived usefulness of mobile learning will have better mobile learning satisfaction. Based on previous suggestions, consequently, this study offers the following hypothesis.

H2: Perceived usefulness could have a positive influence on mobile English learning satisfaction.

Perceived playfulness (PP)

How to make learning be less tedious and more enjoyable to students is always one of the key issues in the educational fields. Although there are several ways to improve learning processes and outcomes, it has been suggested that the applications of new learning technology in instructions play a key role in maximizing learning effectiveness and efficiency (Kopcha, 2010). Previous information technology (IT) research has indicated that user satisfaction is very likely to fall under the sway of perceived playfulness of IT products or services (Hsu & Chiu 2004; Kang & Lee, 2010). Moon and Kim (2001) defined the perceived playfulness of world-wide-web (WWW) as “the extent to which the individual (a) perceives that his or her attention is focused on the interaction with the WWW; (b) is curious during the interaction; and (c) finds the interaction intrinsically enjoyable or interesting” (p. 219). Lee, Yoon, and Lee (2009) added that “playfulness is a complex variable which includes individual’s pleasure, psychological stimulation, and interests” (p. 1323).
Although there is no universal definition of perceived playfulness (Mitchell, Chen, & Macredie, 2005), in mobile learning environments, perceived playfulness, which is regarded as intrinsic motivation (Moon & Kim, 2001), is described as the degree to which a user feels his or her enjoyment, joyfulness, and pleasure in using mobile learning devices to acquire new knowledge (Wang et al., 2009). Several studies have demonstrated a positive link between perceived playfulness and customer satisfaction (Hsu & Chiu, 2004; Kang & Lee, 2010). Nevertheless, a recent study by Kim (2010) has indicated a negative result, which reveals that perceived playfulness could not be associated with customer satisfaction. In mobile learning domains, it is possible that learners with higher perceived playfulness could have better mobile English learning satisfaction. Consequently, this study offers the following hypothesis.

H3: Perceived playfulness could have a positive influence on mobile English learning satisfaction.

### Resistance to change (RTC)

Previous reports have indicated that individual resistance to change (RTC), which is viewed as a de-motivator (Baddoo & Hall, 2003), is another key element that could negatively affect the use of information technology (Kim & Kankanhalli, 2009; Manzoni & Angehrn, 1997; Nov & Ye, 2008). Especially in online banking studies, it has been shown that whether customers are used to or interested in online banking activities could be very critical for the success in online banking services, mainly because they may not get used to the new way for banking activities (Al-Somali, Gholami, & Clegg, 2009). In mobile learning environments, resistance to change, which is described as learners’ resistance to change from traditional learning ways, is very likely to have a negative impact on mobile technology usages, which in turn could negatively affect mobile learning satisfaction. That is, if users are not used to or interested in using mobile technology to acquire new knowledge, it is possible that they will have lower mobile learning satisfaction. Hence, this study proffers the following hypothesis.

H4: Resistance to change could have a negative influence on mobile English learning satisfaction.

### The moderating effect of self-management of learning

The self-management of learning (SML) has been one of the central issues in educational research (Lounsbury, Levy, Park, Gibson, & Smith, 2009), probably owing to its critical role in facilitating more positive learning performances (Abar & Loken, 2010; Lounsbury et al., 2009; Moos, 2010). Zimmerman and Pons (1986) have suggested that SML capabilities could be one of the key indicators to determine learning achievement. Similar words, close to the meaning of SML, cloud include autonomous learning, independent learning, and self-directed learning (Regan, 2003; Wang et al., 2009). Abar and Loken (2010) indicated that SML “involves activating and sustaining cognitions, behaviors, and emotions in a systematic way to attain learning goals” (p. 25). Zou and Zhang (2013) revealed that “in general, students are self-regulated when they are meta-cognitively, motivationally, and behaviorally active participants in their own learning process, without relying on teachers, parents, or other educational services” (p. 55). Pintrich (1999) suggested that SML models could contain “three general categories of strategies: (1) cognitive learning strategies, (2) self-regulatory strategies to control cognition, and (3) resource management strategies” (p. 460).

Although several studies have examined the nature and essence of SML, generally speaking, the historical development, definitions, and strategies of SML could be closely linked to three critical and cyclical stages: goal setting and strategic planning, performance monitoring, and performance management (Zimmerman, Bonner, & Kovach, 1996; Zimmerman, 2000). For example, in order to maximize the effectiveness and efficiency of SML, an early review by Nückles, Hübner, and Renkl (2009) has revealed that writing learning protocol could be a critical way to enhance SML outcomes. Another recent report by Kostons, Gog, and Paas (2012) has shown that students’ SML could be facilitated by training their self-assessment and task-selection skills.

With particular respect to the association between SML, extrinsic, and intrinsic motivation, Sha, Looi, Chen, Seow, and Wong (2012) have not only suggested that according to the self-determination theory (SDT) (Ryan & Deci, 2000), the need for autonomy, which is regarded as intrinsic motivation, is critical to the success of mobile learning, but also added that “the variance of student performance and achievement in mobile learning can be accounted for by the degree to which individual students are motivated intrinsically to ubiquitously engage in mobile learning activities” (p. 720). Additionally, in an adult learning study, Ahmad and Majid (2010) have indicated that self-regulated learners tend to be more intrinsically motivated and be fond of learning. In another recent report, Moos (2010) has further suggested that the levels of quality in self-management of learning could be closely linked to
intrinsic and extrinsic motivation. That is, in self-management of learning, it is revealed that deep learning could be associated with intrinsic motivation, whereas surface learning could be connected with extrinsic motivation (Moos, 2010).

In terms of the moderating role of SML in mobile learning, it is likely that learners with better SML will have a stronger relationship between perceived playfulness and MELS, a weaker connection resistance to change and MELS, and a stronger relationship between MELS and MELCI than those with less SML, due probably to the critical roles of intrinsic motivation variables, such as perceived playfulness and satisfaction, in learning outcomes (Ahmad & Majid, 2010). Conversely, it is assumed that learners with less SML have a better link between perceived usefulness and MELS than those with better SML, probably because they could be more extrinsically motivated (Ahmad & Majid, 2010). In other words, learners with less SML are likely to rely more on the usefulness of mobile devices, which in turn could lead to better mobile learning satisfaction (Arbaugh, 2000; Davis, 1989; Roca et al., 2006).

In mobile learning environments, Wang et al. (2009) described SML as “the extent to which an individual feels he or she is self-disciplined and can engage in autonomous learning” (p. 101). That is, as learners could spend more time in their autonomous and independent learning activities with specific regard to learning goals, it is possible that they will have better learning achievements. Considering the critical impacts of SML on learning outcomes, although numerous researchers have focused on the relationship between SML and learning achievements, little is known about the moderating role of SML in mobile learning satisfaction and continuance intention. Based on previous suggestions, it is conceivable that the self-management of learning could moderate the relationships between key mobile learning determinants, satisfaction and continuance intention. Accordingly, this study proposes the following hypotheses.

H5: The self-management of learning could moderate the relationship between perceived usefulness and MELS. That is, learners with less SML could have a stronger relationship between perceived usefulness and MELS than those with better SML.

H6: The self-management of learning could moderate the relationship between perceived playfulness and MELS. That is, learners with better SML could have a stronger relationship between perceived playfulness and MELS than those with less SML.

H7: The self-management of learning could moderate the relationship between resistance to change and MELS. That is, learners with better SML could have a weaker relationship between resistance to change and MELS than those with less SML.

H8: The self-management of learning could moderate the relationship between MELS and MELCI. That is, learners with better SML could have a stronger relationship between MELS and MELCI than those with less SML.

In summary, the primary purpose of this study is to explore the moderating role of self-management of learning in mobile English Learning. According to previous suggestions, consequently, this study proposes the following research framework (see figure 1).

![Figure 1. The research framework](image-url)
Research methodology

Demographic data for respondents

The participants of this study were 389 undergraduate students in Taiwan. As shown in the table 1, there were more female than male students participating in this study (Male = 170; Female = 215). With regard to the academic level of participants, the number of freshman, sophomore, junior, and senior students was 132, 146, 53, and 46, respectively. It was revealed that most participants were undergraduate students majoring in business (see table 1).

<table>
<thead>
<tr>
<th>Table 1. Demographic data for respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Undergraduate Level</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>College Major</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Data collection

This study gathered the data from 9 different universities, which were randomly selected from 4-year colleges and universities in Northern, Middle, and Southern Taiwan. 900 paper and pencil surveys were sent to the participants. In addition, the participation of this study was voluntary, and all participants were undergraduate students that used to use handheld electronic dictionaries to learn English before. The final number of usable data was 389, after this study deleted the invalid surveys.

Measurement development

18 questionnaire items were scored on a seven-point Likert scale, with rank from “1 = strongly disagree” to “7 = strongly agree”. Three items of the perceived playfulness were selected from Ahn, Ryu, and Han (2007), and from Igbaria, Livari and Maragath (1995). Four items of perceived usefulness were chosen from Davis (1989), and Roca et al. (2006). Four items of resistance to change were taken from Al-Somali et al. (2009), and four items of self-management of learning were adopted from Wang et al. (2009). Moreover, three items of mobile English learning satisfaction, and three items of mobile English learning continuance intention were adopted from Roca et al. (2006).

Common method bias

In order to evaluate the potential threats related to common method bias, Harman’s single-factor test was adopted to investigate the effect of common method bias on this study (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). It was shown that one factor only accounted for 37.4% of total variances. Accordingly, the common method bias may not be a major concern for this study.

Control variables

The types of college and universities, and user experience were two major control variables in this study. That is, this study only collected data from 4-year colleges and universities in Taiwan, and the participants of this study were undergraduate students that used to use handheld electronic dictionaries to learn English before.
Data analysis and results

The Partial Least Squares (PLS) analysis, a component-based Structural Equation Modelling (SEM) technique (Chin, Marcolin, & Newsted, 2003), was adopted to examine the data in this study. First, the factor loadings and Composite Reliability (CR) in each model were used to determine the internal consistency and reliability of measuring scales. As shown in table 2, the internal consistency and reliability of measuring scales were acceptable, mainly because factor loadings in each model were all above the suggested value = .70, and Composite Reliability of each construct was all above .90 (Fornell, & Larcker, 1981). In addition, as shown in table 3, it was revealed that the convergent and discriminant validity were acceptable, because the average variance extracted (AVE) in each model was all above the suggested value of .50 (Fornell & Larcker, 1981), and correlations between variables in each model were all lower than the square root of AVE values on the diagonal (Fornell & Larcker, 1981). Third, the structural model and hypotheses were examined through path coefficients in each model. In figure 2, 3, and 4, it was demonstrated that H2 and H3 were supported by study findings, whereas H4 was partly supported by study results in the full model and low SML group.

Finally, in order to investigate the moderating effect of self-management of learning on mobile learning, the median value of SML = 4.25 was adopted to classify participants into two groups: high SML group \(n = 194\), and low SML group \(n = 195\). According to procedures from Keil, Tan, Wei, and Saarinen (2000), the analysis of path coefficient comparison, which was initially suggested by Wynne Chin and subsequently adopted by several studies (Chung & Kwon, 2009; Fang, 2012; Hartmann & Slapničar, 2012; Hwang, 2010; Kwahk & Ahn, 2010; Lee, Shi, Cheung, Lim, & Sia, 2011; Sanchez-Franco, Ramos, & Velicia, 2009), was performed to probe into the moderating role of self-management of learning in mobile learning. The procedures were as follows:

\[
\text{Spooled} = \sqrt{\frac{(Nh - 1) \times SEh^2 + (Nl - 1) \times SEL^2}{Nh + Nl - 2}}
\]

\[
t = \frac{(PCh - PCI)}{\text{Spooled}} \sqrt{\frac{1}{Nh} + \frac{1}{Nl}}
\]

Spooled = pooled estimator for the variance
\(t\) = t-statistic with \((Nh+Nl-2)\) degrees of freedom
\(Nh\) = sample size of High SML group; \(Nl\) = sample size of Low SML group
\(PCh\) = path coefficient in structural model of high SML
\(PCI\) = path coefficient in structural model of low SML
\(SEh\) = standard error of path in structural model for high SML
\(SEl\) = standard error of path in structural model for low SML

As shown in table 4, it was found that H5, H6, H7, and H8 were all supported by study findings, which indicated that the self-management of learning could moderate the relationships between key mobile learning determinants, satisfaction and continuance intention.
Table 2. Confirmatory factor analysis of each model

<table>
<thead>
<tr>
<th>Items</th>
<th>FM</th>
<th>CR</th>
<th>AVE</th>
<th>α</th>
<th>High SML</th>
<th>CR</th>
<th>AVE</th>
<th>α</th>
<th>Low SML</th>
<th>CR</th>
<th>AVE</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>.92</td>
<td>.74</td>
<td>.88</td>
<td>.84</td>
<td>.92</td>
<td>.74</td>
<td>.88</td>
<td>.84</td>
<td>.92</td>
<td>.73</td>
<td>.88</td>
<td>.88</td>
</tr>
<tr>
<td>PU1</td>
<td>.84</td>
<td></td>
<td></td>
<td>.90</td>
<td></td>
<td>.89</td>
<td></td>
<td>.87</td>
<td></td>
<td>.92</td>
<td>.79</td>
<td>.87</td>
</tr>
<tr>
<td>PU2</td>
<td>.86</td>
<td></td>
<td></td>
<td>.85</td>
<td></td>
<td>.85</td>
<td></td>
<td>.85</td>
<td></td>
<td>.92</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>PU3</td>
<td>.89</td>
<td></td>
<td></td>
<td>.90</td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.91</td>
<td>.72</td>
<td>.87</td>
</tr>
<tr>
<td>PU4</td>
<td>.92</td>
<td></td>
<td></td>
<td>.91</td>
<td></td>
<td>.73</td>
<td></td>
<td>.88</td>
<td></td>
<td>.91</td>
<td>.72</td>
<td>.87</td>
</tr>
<tr>
<td>PP</td>
<td>.93</td>
<td>.82</td>
<td>.89</td>
<td>.89</td>
<td>.92</td>
<td>.80</td>
<td>.87</td>
<td>.92</td>
<td>.92</td>
<td>.79</td>
<td>.88</td>
<td>.88</td>
</tr>
<tr>
<td>PP1</td>
<td>.89</td>
<td></td>
<td></td>
<td>.85</td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>PP2</td>
<td>.92</td>
<td></td>
<td></td>
<td>.79</td>
<td></td>
<td>.79</td>
<td></td>
<td>.79</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>PP3</td>
<td>.92</td>
<td></td>
<td></td>
<td>.83</td>
<td></td>
<td>.83</td>
<td></td>
<td>.83</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>RTC</td>
<td>.88</td>
<td></td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>RTC1</td>
<td>.88</td>
<td></td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>RTC2</td>
<td>.94</td>
<td></td>
<td></td>
<td>.90</td>
<td></td>
<td>.90</td>
<td></td>
<td>.90</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>RTC3</td>
<td>.81</td>
<td></td>
<td></td>
<td>.81</td>
<td></td>
<td>.81</td>
<td></td>
<td>.81</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>RTC4</td>
<td>.80</td>
<td></td>
<td></td>
<td>.80</td>
<td></td>
<td>.80</td>
<td></td>
<td>.80</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>MELS</td>
<td>.82</td>
<td></td>
<td></td>
<td>.82</td>
<td></td>
<td>.82</td>
<td></td>
<td>.82</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>MELS1</td>
<td>.92</td>
<td></td>
<td></td>
<td>.92</td>
<td></td>
<td>.92</td>
<td></td>
<td>.92</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>MELS2</td>
<td>.89</td>
<td></td>
<td></td>
<td>.89</td>
<td></td>
<td>.89</td>
<td></td>
<td>.89</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>MELS3</td>
<td>.92</td>
<td></td>
<td></td>
<td>.92</td>
<td></td>
<td>.92</td>
<td></td>
<td>.92</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>MELCI</td>
<td>.87</td>
<td></td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.87</td>
<td></td>
<td>.90</td>
<td>.75</td>
<td>.83</td>
</tr>
</tbody>
</table>
MELCI1 .90 .92 .88
MELCI2 .91 .93 .87
MELCI3 .86 .83 .87

Notes. FM, full model; High SML, high self-management of learning; Low SML, low self-management of learning; PU, perceived usefulness; PP, perceived playfulness; RTC, resistance to change; MELS, mobile English learning satisfaction; MELCI, mobile English learning continuance intention; CR, Composite Reliability; AVE, Average Variance Extracted; α, Cronbach’s Alpha.

Table 3. The Correlations of each latent variable among different models

<table>
<thead>
<tr>
<th>Full Model</th>
<th>PU</th>
<th>PP</th>
<th>RTC</th>
<th>MELS</th>
<th>MELCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Playfulness (PP)</td>
<td>.44</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resistance to Change (RTC)</td>
<td>-.35</td>
<td>-.18</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile English Learning Satisfaction (MELS)</td>
<td>.56</td>
<td>.49</td>
<td>-.31</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>Mobile English Learning Continuance Intention (MELCI)</td>
<td>.58</td>
<td>.45</td>
<td>-.41</td>
<td>.65</td>
<td>.88</td>
</tr>
<tr>
<td>High SML</td>
<td>PU</td>
<td>PP</td>
<td>RTC</td>
<td>MELS</td>
<td>MELCI</td>
</tr>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Playfulness (PP)</td>
<td>.46</td>
<td>.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resistance to Change (RTC)</td>
<td>-.36</td>
<td>-.15</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile English Learning Satisfaction (MELS)</td>
<td>.49</td>
<td>.55</td>
<td>-.26</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Mobile English Learning Continuance Intention (MELCI)</td>
<td>.59</td>
<td>.52</td>
<td>-.34</td>
<td>.67</td>
<td>.89</td>
</tr>
<tr>
<td>Low SML</td>
<td>PU</td>
<td>PP</td>
<td>RTC</td>
<td>MELS</td>
<td>MELCI</td>
</tr>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Playfulness (PP)</td>
<td>.32</td>
<td>.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resistance to Change (RTC)</td>
<td>-.41</td>
<td>-.31</td>
<td>.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile English Learning Satisfaction (MELS)</td>
<td>.59</td>
<td>.38</td>
<td>-.43</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>Mobile English Learning Continuance Intention (MELCI)</td>
<td>.53</td>
<td>.28</td>
<td>-.58</td>
<td>.61</td>
<td>.87</td>
</tr>
</tbody>
</table>

Notes. Diagonal elements are the square root of Average Variance Extracted

Table 4. Statistical comparison of moderating effect models

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Path</th>
<th>High SML (N=194)</th>
<th>Low SML (N=195)</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Path coefficient</td>
<td>Standard Error</td>
<td>Path coefficient</td>
</tr>
<tr>
<td>H5</td>
<td>PU→MELS</td>
<td>.270</td>
<td>.062</td>
<td>.459</td>
</tr>
<tr>
<td>H6</td>
<td>PP→MELS</td>
<td>.411</td>
<td>.048</td>
<td>.173</td>
</tr>
<tr>
<td>H7</td>
<td>RTC→MELS</td>
<td>-.108</td>
<td>.060</td>
<td>-.192</td>
</tr>
<tr>
<td>H8</td>
<td>MELS→MELCI</td>
<td>.673</td>
<td>.039</td>
<td>.612</td>
</tr>
</tbody>
</table>

Notes. PU, perceived usefulness; PP, perceived playfulness; RTC, resistance to change; MELS, mobile English learning satisfaction; MELCI, mobile English learning continuance intention.

Discussions and implications

In accordance with previous research, the study results have shown that perceived usefulness (Arbaugh, 2000; Roca et al., 2006; Sun et al., 2008), playfulness (Hsu & Chiu, 2004; Kang et al., 2009; Kang & Lee, 2010), and resistance to change could be closely related to mobile learning satisfaction (Al-Somali et al., 2009). Additionally, it has been demonstrated that the study findings are in line with previous suggestions (Abar & Loken, 2010; Zou & Zhang, 2013; Moos, 2010), which indicate that self-management of learning could moderate the relationships between key mobile learning determinants, satisfaction, and continuance intention (see table 4).
With specific regard to learners with higher SML, it has been found that playfulness of mobile devices, an intrinsic motivation, could play the most important role in determining mobile learning satisfaction, probably because they are more intrinsically motivated by the playfulness of mobile devices (Moos, 2010). Conversely, it has been revealed that learners with lower SML could focus more on usefulness of mobile devices, maybe because the useful functions of mobile devices, an extrinsic motivation, could be more helpful and valuable to them especially in increasing mobile learning performance (Davis 1989; Moos, 2010; Roca et al., 2006). It is implied that if instructors would like to help mobile learners with higher SML, intrinsic motivators could play key roles in facilitating them to have more positive learning achievement and deep learning, whereas with respect to mobile learners with lower SML, extrinsic motivators could be more suitable for them in order to help them enhance their learning outcome (Moos, 2010). Accordingly, in order to facilitate learners to achieve better mobile learning performance, it is important that more attention should be given to mobile learning designs that address learners with different SML capabilities.

Moreover, it has been demonstrated that learners with higher SML could have a stronger relationship between mobile learning satisfaction and continuance intention than those with less SML. The study findings, consistent with previous reports (Abar & Loken, 2010; Zou & Zhang, 2013), have further indicated that SML could play a moderating role in determining mobile learning outcomes. In order to improve mobile learning outcome, it is critical that researchers and practitioners should focus more on learners’ SML capabilities than on mobile technology adoption, and more efforts should be directed toward instructional strategies which could improve students’ SML capabilities. For example, motivating students to write learning protocol or facilitating them to get further trainings in self-assessment and task-selection skills could be key ways to help learners with less SML enhance their mobile learning achievements (Kostons, Gog, & Paas, 2012; Nückles, Hübner, & Renkl, 2009).

Last but not least, in terms of the effect of resistance to change on mobile learning, the study results are partially consistent with previous reports (Kim & Kankanhalli, 2009; Manzoni & Angehrn, 1997; Nov & Ye, 2008), which reveal that with particular respect to learners with higher SML, resistance to change could have no influence on mobile learning satisfaction. It is implied that more work should be done on minimizing the resistance to change especially for learners with lower SML, mainly because they could be more reluctant to take mobile learning, which in turn could have a negative impact on mobile learning outcome.

**Limitations and conclusions**

First, with respect to the generalization and extrapolation of study results, it is important that study findings should be interpreted with caution, mainly because of limited resources available for data analysis. Moreover, it is suggested that the moderating roles of age differences and user experience, which could be closely linked to the success of mobile learning, should be further examined in future studies (Lin, 2011). In conclusion, the study results have not only added to the body of knowledge in the educational technology and mobile learning field, but also provided researchers and practitioners with useful information to improve mobile learning designs. As mobile learning has gradually become more indispensable in our lives, the self-management of learning will play a more important role in lifelong learning. Thus, it is necessary that more attention and research should be devoted to the moderating effect of self-management of learning on mobile learning outcomes.

**References**


### Appendix

Mobile English learning questionnaire

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived Usefulness</strong></td>
<td>PU1. Using electronic dictionary to learn English could improve my English learning performance. PU2. Using electronic dictionary to learn English could enhance my English learning effectiveness. PU3. Using electronic dictionary to learn English could make English learning easier. PU4. I found the electronic dictionary to be useful to me in my English learning.</td>
</tr>
<tr>
<td><strong>Perceived Playfulness</strong></td>
<td>PP1. Using electronic dictionary to learn English is one of my enjoyments. PP2. Using electronic dictionary to learn English gives learning fun to me. PP3. Using electronic dictionary to learn English is pleasurable to me.</td>
</tr>
<tr>
<td><strong>Resistance to Change</strong></td>
<td>RTC1. I am not interested in new mobile learning technological developments. RTC2. I feel uncomfortable in changing my current learning methods and using electronic dictionary to learn English. RTC3. I am not interested to use electronic dictionary to learn English. RTC4. I am not used to using electronic dictionary to learn English.</td>
</tr>
<tr>
<td><strong>Self-Management of Learning</strong></td>
<td>SML1. When it comes to learning and studying, I am a self-directed person. SML2. In my studies, I am self-disciplined and find it easy to set aside reading and homework time. SML3. I am able to manage my study time effectively and easily complete assignments on time. SML4. In my studies, I set goals and have a high degree of initiative.</td>
</tr>
<tr>
<td><strong>Mobile English Learning Satisfaction</strong></td>
<td>MELS1. I am satisfied with my electronic dictionary. MELS2. I feel that using electronic dictionary serves my need for learning English very well. MELS3. My decision to use electronic dictionary to learn English is a wise one.</td>
</tr>
<tr>
<td><strong>Mobile English Learning Continuance Intention</strong></td>
<td>MELCI1. I will continue to use electronic dictionary to learn English in the future. MELCI2. I intend to regularly use electronic dictionary to learn English. MELCI3. I would recommend to other students to use electronic dictionary to learn English.</td>
</tr>
</tbody>
</table>
Development of the Interactive Whiteboard Acceptance Scale (IWBAS): An Initial Study

Kung-Teck Wong¹*, Timothy Teo² and Pauline Swee Choo Goh³

¹Sultan Idris, Education University, Malaysia // ²University of Macau, Macau SAR, China // ³Sultan Idris, Education University, Malaysia // Thomas@fppm.upsi.edu.my // timothyteo@umac.co // goh.sc@fppm.upsi.edu.my

*Corresponding author

(Submitted July 15, 2013; Revised January 20, 2014; Accepted April 3, 2014)

ABSTRACT

The purposes of this study were to develop and to conduct an initial psychometric evaluation of the Interactive Whiteboard Acceptance Scale (IWBAS). The process of item-generation for the IWBAS was carried out through the sequential mixed-method approach. A total of 149 student teachers from a teacher-education institution in Australia participated in the project. The principal component analysis (PCA) yielded a five-factor model comprising 14 items, and its factorial validity was confirmed through the use of confirmatory factor analysis (CFA) using structural equation modelling (SEM). The IWBAS reached the minimum thresholds for an acceptable model fit. In addition to the factorial validities, the convergent validity and discriminant validity of the IWBAS were examined, both showing satisfactory validity and good internal consistency for all five constructs. On this basis, the IWBAS can be considered a valid and reliable instrument designed specifically for assessing IWB acceptance among student teachers.

Keywords
Interactive whiteboard, Student teachers, Educational technology, Technology acceptance

Introduction

The introduction of the interactive whiteboard (IWB) into the multimedia learning environment has ushered in a new era of educational technology. It is a versatile, multi-touch, multi-user interactive learning board that allows groups of learners to work simultaneously. First introduced by SMART Technologies, it found ready acceptance in the United Kingdom for office presentations, and IWBs soon became a mainstay in the British education system (Higgins, Beauchamp, & Miller, 2007). From the research literature it is evident that the use of the boards has led to improved teaching and learning in schools (Hennessy, Deaney, Ruthven, & Winterbottom, 2007; Higgins et al., 2007; Murcia & Sheffield, 2010; Preston & Mowbray, 2008; Wong, Goh, & Osman, 2013). The IWB is a stand-alone interactive display that runs on the teacher’s computer or laptop. IWBs can be placed in any suitable position such as mounted on a wall or on a floor stand.

Advocates have noted that one of the most obvious distinctions between IWB technology and other technologies incorporating a data projector and a computer is that users are able to control the computer at the touch of the whiteboard. It also has the advantage that teachers can remain in the front of the class and still be interacting with both the students and the computer applications. As Wong, Goh, & Osman (2013) have noted, this enables teachers to explain topics and teach with more focus. In addition, teachers and students can seamlessly alternate between texts, sound, video and other applications through interacting with the whiteboard whenever required. With the multi-modality and versatile features of the IWB it is not surprising that teaching and learning have become more interesting, fun, and creative. Reports by Becta, an advisory body for educational technologies in British schools, has identified four advantages for students from the use of IWBs; increased enjoyment and motivation, greater opportunities for participation and collaboration, decreased need for note-taking through the capacity to print from the screen, and the potential to cater for different learning styles (Becta, 2004). Evidence reported in recent literature shows that the IWB is no longer a peripheral add-on feature of the computer but is currently considered an integral part of the instructional process (Campbell & Kent, 2010; Northcote, Mildenhall, Marshall, & Swan, 2010; Türel, Serenko, & Bontis, 2011; Wong, Russo, & McDowall, 2013). White (2007) has estimated that, globally, about 750,000 IWBs have been installed in schools across all levels. At the country level, Türel (2011) calculated that about 50 percent of classrooms in Denmark, 47 percent in The Netherlands, and 45 percent in Australia have IWBs installed. In Australia, for example, studies conducted at Abbotsleigh Junior School over the past five years have concluded that IWBs have enhanced teaching and learning as well as facilitating more collaborative and constructive learning (Preston & Mowbray, 2008). Researchers have attributed the success of the IWB for teaching and learning.
to its alignment with the views of Piaget (1967) and Vygotsky (1978) on “constructivist” teaching. Constructivist learning theory contends that learners create knowledge through personal experiences. Using the board, students are able to compile diagrams and maps based on their prior knowledge and in so doing they initiate active discussions and encourage constructivist learning environments. Furthermore, the IWB has a large-screen display and promotes task authenticity, especially when teachers connect it to online news sites, so bringing the world into the classroom and encouraging further constructive discussions (Wong, Goh, & Osman, 2013). Similarly, students can participate actively in lessons in which they explore problems and learn through hands-on and collaborative experiences.

In response to the growing popularity of the IWB, many researchers have begun to assess its affordability and to evaluate its contributions to subject-specific forms of instruction, such as in mathematics (Jewitt, Moss, & Cardini, 2007; Swan & Marshall, 2010), science (Hennessy et al., 2007; Murcia & Sheffield, 2010; Wong, Goh, & Osman, 2013), teaching in elementary schools (Goodwin, 2008); literacy (Shenton & Pagett, 2007); writing (Martin, 2007); early childhood (Wong, Russo, & McDowall, 2013) and special needs (Keay-Bright, Norris & Owen, 2007). The findings of these various studies attest, either directly or indirectly, to the status of IWB as a new-generation classroom tool which has the capability to encourage the engagement of digital learners in teaching and learning in ways that are more effective than other teaching technologies. With its distinctive features, the IWB has been found to be a suitable tool that can seamlessly engage with conventional pedagogy practices (Betcher & Lee, 2009).

The study

To date, despite the credit and adulation given to IWBs for their effectiveness in teaching and learning, there has been a paucity of studies exploring and examining IWB acceptance among student teachers. It is noteworthy that despite the numerous technology-acceptance measurements that have been formulated to understand acceptance of different types technology in different settings (Ajzen & Fishbein, 1980; Davis, Bagozzi, & Warshaw, 1989; Davis, 1989; Venkatesh, Morris, Davis, & Davis, 2003; Wong, Teo & Russo, 2013), none of the measurements has been particularly pertinent to the acceptance of IWB in teaching and learning.

Some recent studies (Türel, 2011; Türel & Johnson, 2012; Şad, 2012) have aimed to empirically develop a valid and reliable IWB measurement to explore IWB use by researching the perceptions of school students or practicing teachers, however, none has focused on student teachers’ perspectives of IWB acceptance. Researchers believe that student teachers have different backgrounds, mindsets and attitudes regarding IWB acceptance. It is noteworthy that school students and student-teachers in teacher education institutions have quite different reasons for using IWBs (Wong, Russo, & McDowall, 2013). Furthermore, the type of technology selected for teaching and learning practice by student teachers is based mainly on the topic and its objectives. In addition, student teachers tend to have freedom to use any teaching tools which they consider best suited to their planned lessons (Teo, Wong, & Chai, 2008). Along the same lines, advocates of IWBs have highlighted that user groups and applications engaged with the technology were significant predictors for technology acceptance in various settings (Im, Kim, & Han, 2008; Marchewka & Kostiwa, 2007; Venkatesh et al., 2003). As a consequence, it was deemed more accurate to measure technology acceptance based on technology type, its applications, and the group involved.

On this basis, having a valid and reliable instrument designed specifically for assessing IWB acceptance by student teachers could provide insights into issues relating to its use. Furthermore, such information would be useful for policymakers and teacher-educators, especially in the design of curricula which could enhance learning experiences for trainee teachers, thus encouraging them to engage in using new technologies in their future teaching and learning activities.

Given the reported effectiveness and versatility of the IWB in education, this researcher’s attention has been drawn to consider the level of acceptance of the IWB by student teachers. This interest is consistent with evidence to suggest that student teachers’ opinions may mirror those of future teachers (Teo, Lee, & Chai, 2008; Wong, Teo, & Russo, 2012). In addition, so far a literature review has not yielded any references to instruments developed to understand student-teacher acceptance of the IWB. It is pertinent to note at this point that over the past three or four decades new technologies have not always been readily adopted by educators and neither have they been used as manufacturers have expected. Consequently, a valid and reliable instrument designed for assessing IWB acceptance among student teachers could provide insights into issues relating to the use of technology in schools and other educational sectors. It is reasonable to believe that a valid IWBAS could enhance the understanding of teacher-
educators and related policymakers regarding the factors influencing the acceptance of IWB by student teachers, thus ensuring that trainees engage IWBs in their future teaching and learning activities. Thus, having a specific measurement or scale to evaluate IWB acceptance among student teachers is a worthwhile issue for serious enquiry.

The aim of this study was to develop a valid and reliable Interactive Whiteboard Acceptance Scale (IWBAS) in order to explore and understand factors influencing student-teacher acceptance of the IWB.

Method

Item generation

The process of item generation for the IWBAS was carried out through the sequential exploratory mixed-method, a powerful method for developing and validating a new instrument (Creswell, 2003). In the current study the researchers commenced with a qualitative approach in order to establish the content and validity of the scale, this being followed by quantitative analyses to assess the factor structure and psychometric properties. The exploratory mixed-method has been used in previous studies for developing and validating measurements in educational settings (Teo, 2010; Türel, 2011; Şad, 2012). In the qualitative phase, a review of the literature of similar fields was carried out, with particular emphasis on the items employed in empirical studies (Compeau & Higgins, 1995; Davis, 1989; Gibson & Dembo, 1984; Riggs & Enochs, 1990; Teo, 2009; Thompson, Higgins, & Howell, 1991; Venkatesh et al., 2003).

The literature relating to this area provided a practical framework for constructing the IWBAS (Teo, 2010; Türel, 2011; Şad, 2012; Venkatesh et. al., 2003). This was supplemented by information obtained when the researchers interviewed seven student teachers in order to help build the item pool. The interviews consisted of predetermined questions derived from the literature on similar fields of enquiry. The participating student teachers were recommended by colleagues and through personal contacts. A total of 28 items were created in the first draft. Next, before an examination of the psychometric quality of the IWBAS scale was performed, verification of the content and criterion-related validity was carried out. The researchers consulted several academic several academic researchers experienced in instructional design, three IWB certified trainers, and five teachers experienced in the use of the IWB, their advice being invaluable in determining the items to be used for nomological validity. Next, the items were given to ten student teachers for preliminary examination. The students were randomly selected (being non-participants in the previous interviews) and were asked to explain to researchers the meaning of each item.

Subsequently, a 14-item scaled questionnaire was compiled to assess the acceptance of the IWB. The IWBAS consisted of a four-point Likert scale with responses ranging from “strongly disagree” (1) to “strongly agree” (4). The four-point scale was employed to minimise both a social desirability bias (Garland, 1991; Worcester & Burns, 1975)and respondents’ tendency to choose a midpoint on a five-point format (Matell & Jacoby, 1972). For the quantitative phase, an examination of the psychometric quality of the IWBAS was carried out to ensure the validity and reliability of the items used. For these purposes, a principal component analysis (PCA) and a confirmatory factor analysis (CFA) were implemented.

Participants and data-collection methods

Invitations to participate in this study were extended to student teachers enrolled in science-related courses including the Bachelor of Early Childhood Education, Bachelor of Education (Junior Primary and Primary), and Bachelor of Education (Primary and Middle). A total of 149 agreed to participate: 112 from the first course, 17 from the second, and 20 from the third. Of these participants 146 (98 percent) were female because of the preponderance of female student teachers in the early- and primary-teacher education programmes. Most of them (88.1 percent) had not attended any formal IWB training or workshop, although 74.8 percent reported having had some previous experience with using IWBs for teaching and learning. Participation was voluntary and anonymous and no course credits were given. All participants were briefed on the purpose of the study and informed of their right to withdraw during or after data collection. Participants took approximately 20 minutes to complete the questionnaire.
Results

Psychometric quality of the instrument

In order to develop a valid and reliable IWBAS, close examinations of the factor structure and psychometric properties of the items were conducted. For these purposes, a principal component analysis (PCA) and a confirmatory factor analysis (CFA) were performed.

Principal component analysis (PCA)

A principal component analysis (PCA) was carried out to examine and compute composite scores for the constructs underlying the IWBAS. It is common to assess the factorial validity of the scale used in a research project, and prior to conducting the principal component analysis it is sound practice to appraise the suitability of the dataset. All 14 items were examined for their mean, standard deviation, skewness, and kurtosis. All mean scores were above the midpoint of 2.5, the range being between 2.8 and 3.2. This indicated an overall positive response to the constructs in the study. The standard deviation (SD) values for all constructs were less than one. This indicated that the item scores had a relatively narrow spread around the mean. The univariate normality of the dataset was assessed through inspection of skewness and kurtosis, with values less than 3.0 and 10 respectively, indicative of acceptable normality (Kline, 2010). From the findings, the skewness (performance expectancy = -.94; effort expectancy = .32; social influence = -.29; facilitating conditions = -.97; self-efficacy = .36) and kurtosis (performance expectancy = .19; effort expectancy=.48; social influence = .30; facilitating conditions = .98; self-efficacy = .17) indicated that all constructs were acceptable. In addition to the normality test, the results of the Kaiser-Meyer-Olkin (KMO) test (.685) and Bartlett’s test of sphericity (BTS) ($\chi^2 = 764.008; df = 91; p < .001$) indicated the dataset was adequate for factorability analysis.

After the dataset was assessed for its factorability, the 14 items of the IWBAS were subjected to factor analysis. An eigenvalue greater than 1 should be achieved to determine the number of components in the scale (Kaiser, 1960). Based on Kaiser’s assumption, the suggested IWBAS had five components with eigenvalues exceeding 1. The results depicted in Table 1 show that the theoretical five-factor structure in the IWBAS explained 70.98 percent of the total variance where the first factor had an eigenvalue of 3.14 and explained 22.44 percent of the total variance while the second factor had an eigenvalue of 2.29 and explained 16.35 percent of the total variance. The third factor had an eigenvalue of 1.78 and explained 12.73 percent of the total variance. The fourth and last factors contributed 19.44 percent to the total variance of the IWBAS. Furthermore, the results indicated that all the factor loadings of the individual items were above .50 and ranged from .68 to .93. None of the items reflected high factor loadings on a second or additional factor. Principal Axis Factoring (PAF) was carried out to cross check the PCA results, and the variances explained from the PAF were similar to the PCA. Consequently, these results confirmed that at least half the variances in all the indicators were explained by their respective latent constructs. Hair, Black, Babin and Anderson (2010) suggested that an item is significant if its factor loading is greater than .50.

Based on the five-factor structure from the PCA, and further examining the theoretical meaningfulness and coherence of the loaded items in each factor, five constructs were identified. They were performance expectancy (PE) (three items), effort expectancy (EF) (three items), social influence (SI) (three items), facilitating conditions (FC) (three items), and self-efficacy (SE) (two items). In the present context, performance expectancy refers to student-teachers’ belief that using IWBs would help them to attain benefits for their teaching and learning practices. Effort expectancy refers to the extent to which student teachers believe that the use of IWBs would require little effort and enable them to be free of distractions. Social influence refers to the extent to which student teachers’ perception that most people who are important to them think they should or should not use the IWB. Facilitating conditions refer to the degree to which individuals believe that organisational and technical infrastructures exist to support them. On the other hand, self-efficacy refers to student teachers’ judgement of their capabilities to teach with the technology (IWB).
Table 1. Results of a principal component analysis of the item scale

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
<th>Loading</th>
<th>Eigenvalues</th>
<th>Variances explained (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Expectancy</td>
<td>PE1</td>
<td>.77</td>
<td>2.35</td>
<td>16.82</td>
</tr>
<tr>
<td>(PE)</td>
<td>PE2</td>
<td>.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PE3</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort Expectancy</td>
<td>EE1</td>
<td>.88</td>
<td>2.30</td>
<td>16.44</td>
</tr>
<tr>
<td>(EE)</td>
<td>EE2</td>
<td>.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EE3</td>
<td>.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Influence (SI)</td>
<td>SI1</td>
<td>.82</td>
<td>2.05</td>
<td>14.71</td>
</tr>
<tr>
<td></td>
<td>SI2</td>
<td>.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI3</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitating Conditions</td>
<td>FC1</td>
<td>.82</td>
<td>1.84</td>
<td>13.20</td>
</tr>
<tr>
<td>(FC)</td>
<td>FC2</td>
<td>.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FC3</td>
<td>.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Efficacy (SE)</td>
<td>SE1</td>
<td>.75</td>
<td>1.37</td>
<td>9.82</td>
</tr>
<tr>
<td></td>
<td>SE2</td>
<td>.82</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Extraction method: Principal Component Analysis (PCA)
Rotation method: Varimax with Kaiser Normalization

**Confirmatory factor analysis (CFA)**

Factorial validity for the five-factor structure of the 14-item scale was confirmed through the use of confirmatory factor analysis (CFA) using structural equation modelling (SEM). A maximum likelihood estimation (MLE), using underlying AMOS software, was carried out for this purpose. Table 2 shows that all parameter estimates were significant at the $p < .05$ level, as indicated by the $t$-value which was greater than 1.96. All standardised estimated weights were above .50 and ranged from .541 to .968: these values were considered appropriate and acceptable (Hair, et al., 2010). Furthermore, the multiple square correlations ($R^2$) of all items ranged from .30 to .94, suggesting that these items were explained by their predictors in a range between 30 percent and 94 percent. In this study, indices of composite reliability were reported in recognition of the problems associated with the use of Cronbach’s Alpha (Teo& Fan, 2013). These appeared to be above the recommended thresholds for each construct and provided support for the significant indicator-factor relationship of the IWBA scale.

As part of the development of the scale several alternative models should be computed to allow comparisons of different conceptualisations of the factor structure of the proposed scale (Teo, 2010). A series of CFAs was performed to test the models in different factor structures, and the model fit was then assessed. First, a null model (M0) indicated that all the factors were uncorrelated. Second, a one-dimensional structure model (M1) was tested. Finally, a five-structural construct model (M2) was assessed for its fit.

To ensure that the measurement model exhibited a good fit, the following five absolute-fit indices were monitored: Ratio of its degree of freedom ($\chi^2/df$); Goodness of Fit (GFI); Comparative Fit Index (CFI); Tucker-Lewis Index
(TLI); and Standardised Root Mean Square Error of Approximation (RMSEA). Absolute-fit indices measure how well the proposed model reproduces the observed data. According to Hair et al. (2010), the values of GFI and CFI should be more than 0.90 and that of the RMSEA smaller than 0.05 in order to be considered a “good fit.” For $\chi^2/df$, a value below 3.0 is considered acceptable. Finally, the TLI value should be greater than 0.90.

Table 2. Results of the measurement model

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
<th>SE</th>
<th>t-Value</th>
<th>$R^2$</th>
<th>Average Variance Extracted$^a$</th>
<th>Composite Reliability$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE</td>
<td>PE1</td>
<td>.753</td>
<td>-</td>
<td>.568</td>
<td>.70</td>
<td>.851</td>
</tr>
<tr>
<td>PE</td>
<td>PE2</td>
<td>.903</td>
<td>5.61</td>
<td>.815</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>PE3</td>
<td>.553</td>
<td>4.34</td>
<td>.306</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>SE1</td>
<td>.706</td>
<td>3.55</td>
<td>.499</td>
<td>.62</td>
<td>.561</td>
</tr>
<tr>
<td>SE</td>
<td>SE2</td>
<td>.552</td>
<td>-</td>
<td>.305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>EE1</td>
<td>.869</td>
<td>-</td>
<td>.755</td>
<td>.64</td>
<td>.738</td>
</tr>
<tr>
<td>EE</td>
<td>EE2</td>
<td>.746</td>
<td>6.67</td>
<td>.556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>EE3</td>
<td>.522</td>
<td>5.59</td>
<td>.272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>SI1</td>
<td>.791</td>
<td>-</td>
<td>.626</td>
<td>.60</td>
<td>.661</td>
</tr>
<tr>
<td>SI</td>
<td>SI2</td>
<td>.593</td>
<td>4.62</td>
<td>.352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>SI3</td>
<td>.541</td>
<td>4.51</td>
<td>.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td>FC1</td>
<td>.728</td>
<td>8.79</td>
<td>.529</td>
<td>.74</td>
<td>.836</td>
</tr>
<tr>
<td>FC</td>
<td>FC2</td>
<td>.968</td>
<td>8.87</td>
<td>.938</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td>FC3</td>
<td>.702</td>
<td>-</td>
<td>.493</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $^a$t-Value (critical ratio) shows whether the parameter is significant at the .05 level. $^b$AVE: Average Variance Extracted = ($\sum \lambda^2$) / ($\sum \lambda^2 + \sum (1 - \lambda^2)$). $^c$Composite Reliability = ($\sum \lambda^2$) / ($\sum \lambda^2 + \sum (1 - \lambda^2)$). *Indicates an acceptance level or validity. - This value was fixed at 1.00 for model identification purposes, hence no t-value was estimated. SE: Standard Estimate

Table 3. Good-of-fit indices and comparison of alternative models

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>GFI</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>$\chi^2$/df</th>
<th>$A\chi^2$(df)sig</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>791.562*</td>
<td>91</td>
<td>.574</td>
<td>.000</td>
<td>.000</td>
<td>.221</td>
<td>8.698</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-factorial model</td>
<td>607.225*</td>
<td>77</td>
<td>.647</td>
<td>.243</td>
<td>.106</td>
<td>.209</td>
<td>7.886</td>
<td>14(1),182.33**</td>
<td>M1 vs M0</td>
</tr>
<tr>
<td>Five-factorial model</td>
<td>108.063*</td>
<td>66</td>
<td>.910</td>
<td>.940</td>
<td>.917</td>
<td>.064</td>
<td>1.637</td>
<td>11(1),499.16**</td>
<td>M2 vs M1</td>
</tr>
</tbody>
</table>

Note. Ratio of the $\chi^2$ statistic to its degree of Freedom ($\chi^2$/df); Goodness of Fit (GFI); Comparative Fit Index (CFI); Tucker-Lewis Index (TLI); and Root Mean Square Error of Approximation (RMSEA); **p < .01; ns = not significant

From Table 3 it can be seen that the M0 and M1 did not reach the minimum thresholds typically requested for an acceptable fit. However, the M2 showed much better fit statistics and attained the minimum thresholds for an acceptable model fit ($\chi^2 = 108.063^*, p < 0.00; \chi^2$/df = 1.637; GFI = .910; CFI = .940; TLI = .917 and RMSEA = 0.064). However, the $\chi^2$ statistic was found to be too sensitive to sample size differences, especially for studies with large samples. Hair et al. (2010) noted that as the sample size increases there is a great tendency for the $\chi^2$ statistic to indicate significant differences.

Convergent and discriminant validities

In addition to the factorial validities, the convergent validity and discriminant validity of the IWBAS were examined through inspection of the average variance extracted (AVE) and discriminant validity. The AVE for each measure was above .50 which suggested that more than half of the variance observed in the items was accounted for by the hypothesised factors. Table 4 shows the discriminant validity of the measure items. Discriminant validity is present
when the variance shared between a construct and any other constructs in the model is less than the variance that constructs share with their indicators (Teo, 2010). If the square roots of the AVEs are greater than the off-diagonal elements in the corresponding rows and columns, it suggests that the given construct is more strongly correlated with its indicators than with the other constructs in the model (Teo, 2009). From Table 4 it can be seen that the values in the matrix diagonals (representing the square roots of the average variance extracted) are greater than the off-diagonal elements in their corresponding rows and columns, suggesting that discriminant validity was present in the five-factor model with 14 items (performance expectancy (PE) (three items), effort expectancy (EF) (three items), social influence (SI) (three items), facilitating conditions (FC) (three items), and self-efficacy (SE) (two items).

Table 4. Discriminant validity for measurement model

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>EE</th>
<th>SI</th>
<th>FC</th>
<th>CTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE</td>
<td>(.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>.065</td>
<td>(.80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>.162</td>
<td>.046</td>
<td>(.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td>.078</td>
<td>.215</td>
<td>.105</td>
<td>(.86)</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>.362</td>
<td>.177</td>
<td>.174</td>
<td>.118</td>
<td>(.78)</td>
</tr>
</tbody>
</table>

Note. Diagonal in parentheses: square root of average variance extracted from observed variables (items); Off-diagonal: correlations between constructs.

Administering the survey and interpreting scores

It took each participant approximately 20 minutes to complete the IWBAS. The survey consisted of 14 items and participants were requested to select their responses from a four-point scale. In order to determine each participant’s level of acceptance of the IWB he/she was requested to select the most appropriate answer from the following options: strongly disagree (1), slightly disagree (2), slight agree (3) to strongly agree (4). The higher the score, the greater the level of IWB acceptance. The 14-item scores ranged from 14 to 56. Participants were informed that no credit would be given, that pseudo-names would be used, and that they had the right to withdraw during or after data collection.

Limitations of the study

Despite the results indicating that the IWBAS is reliable and valid, some primary limitations mirror the need for further investigation. First, self-reporting items were employed to measure the variables for the present study, thus suggesting the possibility of bias in the findings because participants may have given socially desirable responses, especially since one of the researchers was the course coordinator. Second, the population of this study involved only participants from one teacher-education institution and therefore the findings may not adequately reflect the perceptions of the general student-teacher population in regard to IWB acceptance. Associated with this, the use of the same sample for both exploratory and confirmatory factor analyses was not ideal as this gave rise to a potential capitalization on chance in the latter analysis. To ensure validity any future studies should employ a larger sample from different institutions, and from across a broader set of grade levels, including secondary school, to assess the psychometric properties of the IWBAS. Additionally, future studies should also examine the parsimony, interpretability and consistency of the IWBAS by conducting the survey with practising teachers who would be more likely to have experienced new technologies and had considerable autonomy over their teaching tools. Third, as the theoretical five-factor structure in the IWBAS explained only 70.98 percent of the total variance, it is recommended that future research should include other constructs that may determine IWB acceptance and so account for the unexplained variance. It is relevant to note, too, that although findings from the current study achieved good psychometric characteristics and properties overall, the use of the four-point Likert scale should be employed with caution. Future use of a five- or seven-point Likert psychometric scale could be used to enable comparisons with the current study. It could assess and justify any possibility of acquiescence and social desirability bias in the current scale and, if any, provide better linguistic qualifiers for each item.
Conclusions and recommendations

The purposes of this study were to develop and provide an initial psychometric-property evaluation of a scale to assess the level of acceptance of the IWB. This scale provides a better understanding of IWB acceptance among students in teacher-education institutions, and it highlighted student teachers’ perspectives regarding technology usefulness, ease of use, social encouragement, individual beliefs, and facilitation of support for the use of IWBs. A total of 28 items were created in the first draft, these being reviewed for content and criterion validity by expert panels. Through elimination of unapproved items, 14 items concerning the IWB acceptance scale were assessed by principal component analysis and confirmatory factor analysis. The exploratory factor analysis indicated that the items in the IWBAS consisted of a five-factor structure comprising performance expectancy, effort expectancy, social influence, facilitating conditions, and self-efficacy. The results showed that the theoretical five-factor structure in the IWBAS explained 70.98 percent of the total variance.

In addition to examining the psychometric properties of the IWBAS the researchers also demonstrated the use of confirmatory factor analysis to provide a more parsimonious list of items to measure student-teacher acceptance of the IWB. A series of confirmatory factor analyses was performed to allow comparisons of different conceptualisations of the factor structure and to obtain the best model fit. Results indicated that the hypothesised five-factor structure had good standardised loadings, and all the absolute-fit indices were above the recommended thresholds for an acceptable model fit. Furthermore, results from convergent validity and discriminant validity of the IWBAS showed satisfactory reliability and validity of all five constructs. On this basis, it is concluded that the five-factor structure of the IWBAS is reliable and valid as a scale to measure IWB acceptance among student teachers. Indeed, the scales developed and validated in this study indicated the importance of performance expectancy, effort expectancy, social influence, facilitating conditions and self-efficacy for IWB acceptance.

References


Focused Ubiquity: A Purposeful Approach to Providing Students with Laptops

Mike Keppler¹, Spencer C. Weiler²* and Dan Maas³

¹University of Northern Colorado, 403 McKee Hall, Campus Box 103, Greeley, Colorado 80639 // ²University of Northern Colorado, McKee Hall, Campus Box 103, Greeley, CO 80639 // ³University of Northern Colorado, 5776 South Crocker Street, Littleton, CO 80120 // keppler.mcsa@gmail.com // spencer.weiler@unco.edu // dmaas@lps.k12.co.us

*Corresponding author

(Submitted July 17, 2013; Revised December 19, 2013; Accepted April 8, 2014)

ABSTRACT

Laptops have been introduced into classroom across the nation as a way to improve teaching and learning. In 2007 Littleton Public Schools (LPS) introduced a focused approach to providing all students with a laptop at a significantly lower cost to the traditional ubiquitous laptop programs. The purpose of this study was to document the LPS model and measure its impact on instruction and learning. In addition, a cost analysis comparing the LPS model to other approaches at infusing laptops into classrooms was conducted. The LPS is not only cost effective, the quantitative and qualitative data collected in this study suggest that this approach to laptop infusion in the classroom positively impacts instruction and learning. One implication from this study is that school districts, when considering adopting a ubiquitous laptop program, should work to incorporate a focused curricular approach to increase the likelihood of the laptops improving instruction and learning.

Keywords

Laptops, Student achievement, Cost effective

Introduction

Since the late 1970s, people have ruminated on the significant impact technology can have on student learning. In the early 1980s, as technology began to alter the way business was conducted throughout the world, policy makers emphasized the great promise that technology held for transforming public education. An early enthusiasts felt that computers would radically alter the relationship between teacher and student and that computers would open up the world to students to allow them to access information and discover new ways for creative expression (Papert, 1980).

Research related to the use of computers in the classroom has identified a number of positive influences technology can exert on the learning process. The use of computers in the classroom, coupled with clearly stated learning objectives, results in increased student achievement (Ringstaff & Kelley, 2002). Schacter and Fagnano (1999) argued that technology in the classroom positively impacts student learning, including understanding and achievement, and increases affective attributes. According to Pitler, Hubbell, Kuhn, and Malenoski (2007), computers in the classroom create “dynamic learning environments” and bolster efforts to differentiate instruction (p. 2).

Clariana (2009) reported that 31% of surveyed superintendents were implementing some sort of ubiquitous computer program, or one-to-one computing program, that provides a computer to every student in their school districts (pp. 5-6). Technology enhances student learning and the use of ubiquitous computer efforts are on the rise, but little is known about the different one-to-one computing options available to educators. The aim of this study was to document an alternative approach to infusing computers into the classroom, measure the impact this focused approach had on both teaching and learning, and conduct a cost effectiveness analysis of the three approaches.

Literature Review

The purpose of the literature review is to provide an overview of the use of technology in the learning process. The literature review is divided into four sections and these sections explore the optimal practices related to technology and student achievement.
Technology and learning

The push for change in education has opened the door to new ideas while stressing the importance of altering the way teachers teach and students learn. In a letter written to the Members of Congress in November 2010, The Secretary of Education for the United State of America, Arne Duncan, stated that:

*The model for learning described in this plan calls for engaging and empowering personalized learning experiences for learners of all ages. The model stipulates that we focus what and how we teach to match what people need to know and how they learn. It calls for using state-of-the-art technology and Universal Design for Learning (UDL) concepts to enable, motivate, and inspire students to achieve, regardless of background, languages, or disabilities. It calls for ensuring that our professional educators are well connected to the content and resources, data and information, and peers and experts they need to be highly effective. And it calls for leveraging the power of technology to support continuous and lifelong learning. (p. v)*

The ideas outlined in the plan entitled *Transforming American Education: Learning Powered by Technology* (U.S. Department of Education, 2010) highlight the need for increased college completion rates among young people in order to ensure America is able to compete in a global economy. The plan also stressed that technology is at the core of virtually every aspect of daily life and work, and must be leveraged in a way to continuously improve the education system from the time children enter academia until they complete higher level schooling and/or enter the workforce.

Part of the push for using technology in schools has come from business leaders who are vocalizing the desire for employees to have strong computer skills. Wagner (2008) identified many skills students need, including critical thinking, problem solving, mental agility, and adaptability. Employers are looking for individuals with strong research skills, strong information technology (IT) skills, and people who are able to think logically and process a lot of information in an organized manner (Rothstein, 2004, p. 151).

Students are a strong voice for expanding the use of technology in schools. For example, more than 55,000 students responded to the following open-ended question: “Today, you and your fellow students are important users of technology. In the future, you will be the inventors of new technologies. What would you like to see invented that you think will help kids learn in the future” (Chen, 2010, p. 244)? The responses were synthesized into this scenario:

*Every student would use a small, handheld wireless computer that is voice-activated. The computer would offer high-speed access to a kid-friendly Internet, populated with websites that are safe, designed specifically for use by students, with no pop-up ads. Using this device, students would complete most of their in-school work and homework, as well as take online classes both at school and at home. Students would use the small computer to play mathematics-learning games and read interactive e-textbooks. In completing their schoolwork, students would work closely and routinely with an intelligent digital tutor, and tap a knowledge utility to obtain factual answers to questions they pose.* (Chen, 2010, p. 244)

Students desire a technology-rich learning environment, which provides them opportunities to explore, learn, and take responsibility for their progress in school.

Approaches at utilizing technology

In an effort to further student and teacher access to technology, some school districts have embraced the idea of providing one computer to every student. Broadly speaking, one-to-one computing initiatives support the idea of increasing student access to technology in schools (Li, 2010). “It is anticipated that the impact of this kind of initiative on student learning and pedagogical practices is likely to be far more pervasive and permeable than the other large-scale information technology (IT) initiatives implemented in the past” (Li, p. 284). In fact, “One-to-one computing is one of the fastest growing yet most controversial phenomena in American classrooms” (Lei & Zhao, 2008, p. 98). Over the last decade, one-to-one projects have dramatically expanded with major initiatives taking place in at least 33 states across the country (Lei & Zhao). Some states have undertaken expensive projects to provide laptops to all secondary students. Proponents of these efforts, “believe that educationally beneficial uses of
computers will emerge spontaneously from the deployments of laptop computers in ratios of one computer per user” (Weston & Bain, 2010, p. 10). A study conducted in 2004 examined the impact of laptop computers use in schools in France concluded, “the inference drawn from the hypotheses regarding the use of portable computers in schools postulates that use of this tool should enable the development of cognitive and behavioral abilities which in turn lead to a marked improvement in pupil attainment” (Jaillet, 2004, p. 115).

Unlimited access to the Internet and the integration of technology as a tool for learning has been viewed by many as a way to transform instruction and learning in classrooms (Halverson & Smith, 2009-2010). Some advocates also feel that one-to-one technology in schools helps to minimize the extent to which differences exist between students in large and small districts and support learning for students who have access to resources and those who typically are less fortunate (Chen, 2010; Penuel, 2006). Referring again to the 2010 National Education Technology Plan, a key element of the plan is to “Ensure that every student and educator has at least one Internet access device and appropriate software and resources for research, communication, multimedia content creation, and collaboration for use in and out of school” (p. xix).

Maximizing technology

In order to prepare teachers and students for the digital-age, considerations must be made as to whether or not educators have the skill set necessary for utilizing the tools within a technology rich classroom. More often than not, when it comes to technology, teachers are learning in parallel with, or at times from, their students (Gorder, 2008). In the case of some veteran teachers, there may be a hesitation to use technology as a tool for learning due to a lack understanding. “Teachers need to know how and why to use technology in meaningful ways in the learning process for technology integration to work” (Gorder, 2008, p. 64). While access to technology in classrooms has greatly increased over the last decade, a study found that less than 32% of educators reported using technology as part of their daily classroom practices (Gillard & Bailey, 2007). Part of the problem may be due to the focus or lack of time allotted for teacher training in the area of technology. “Most of today’s educational technology training for both preservice and practicing teachers tends to focus on the mechanics of computer operation, drill and practice, and tutorial software with little integration into specific subjects” (Harwell, Gunter, Montgomery, Shelton, & West, 2001, p. 260). According to Christensen, Horn, and Johnson, (2008), “Understanding how schools have spent so much money on computers only to achieve such little gain isn’t so hard. Schools have crammed the computers into the existing teaching and classroom models” (p. 84). Christensen et al. contended that the lack of professional development to train teachers to effectively use laptops in the classroom enabled teachers to just use the computers to support their current teaching practices. In other words, without the professional development, teachers will not change their practice to accommodate the laptops, which then minimizes the impact the technological tool can have in the classroom.

In short, “If technology is to be integrated successfully into classroom instruction, teachers must be able to demonstrate successful mastery of technology use themselves” (Harwell et al., 2001, p. 260). Based on this fact, leaders in the field of education should pay close attention to teacher preparedness for the use of technology in the classroom. Ultimately, teachers’ attitudes and beliefs about technology’s role in the classroom are directly related to the amount of professional development they have received (NCES, 2000).

A review of the literature found that teachers who engage in in-depth professional development activities centered on the utilization of technology were more likely to report feeling prepared to use computers and the Internet for instruction (Penuel, 2006). Part of the process also involves changing teachers’ perceptions about the value of technology in daily instruction. “It is believed that teachers’ receptiveness to technology and a re-conceptualization of its role in teaching and learning can drive the development of new pedagogies and curricula and eventually, bring about new and meaningful learning experiences for students” (Li, 2010, p. 285). The key aspect of this shift is dependent on two variables: (1) a teacher being comfortable with the many facets of any given technology and (2) a teacher’s ability to evaluate and integrate the use of that technology into the curriculum (Keengwe, Onchwari, & Wachira, 2008).

In order to avoid frustration that can arise from technical breakdowns, teachers also value knowledgeable technical support. “In addition to professional development, readily available technical support also appears to be important for laptop programs to succeed” (Penuel, 2006, p. 339). If school districts are to find success with the use of one-to-
one computing as a tool for learning, educators must be able to move beyond teaching basic skills and use technology as a tool to enhance those skills outlined by business leaders to prepare students for the global economy (Wagner, 2008). Teacher-leaders described their vision of the future and postulated that:

Because students will have easy access to information, the education delivery systems of the future will demand intensely individualized learning. The scarcest commodity will be attention, and successful educators will be those who can attract and hold students’ interest while helping students develop the habits of mind and the digital facility they need to process and evaluate relevant information. Teachers who can customize learning experiences and facilitate them in both physical and virtual environments will be highly sought after. (Moore & Berry, 2010, p. 36)

The National Education Technology Plan (2010) supported this vision, emphasizing the importance that, “They (teachers) are connected to resources and expertise that improve their own instructional practices, continually add to their competencies and expertise, and guide them in becoming facilitators and collaborators in their students’ increasingly self-directed learning” (p. 40). Ongoing teacher training and development are vital aspects of successful technology use in the classroom environment.

Profile of school districts

The three school districts included in this study are profiled below. Basic demographic data are presented on all three school districts. The profile for Littleton Public Schools (LPS) includes a summary of the one-to-one computing program that the school district developed to focus the use of technology in the classroom on writing.

LPS

LPS is a suburban school district located in the Southeast metro area of Denver, Colorado. The school district includes 22 schools serving approximately 15,000 students. The district supports a student population in which 20.79% qualify for the Federal Free and Reduced Lunch program, with individual schools in the district ranging from a low of 4% eligibility to a high of 81%. The district has earned Colorado’s highest accreditation rating for documented high achievement and high growth indicators. LPS is the only Denver metro-area district to hold this rating and has earned it each of the three years the rating has been given by the Colorado Department of Education (2012).

The Littleton laptop initiative

The Inspired Writing initiative is a technology integration project conducted throughout LPS to connect writing instruction and one-to-one computing use in classrooms. Research conducted by the District leadership prior to the effort suggested that writing achievement could be expected to improve through the instructional use of computers in classrooms (Pitler, Flynn & Gaddy, 2004). As the Learning Services department engaged in a Literacy Initiative, the Information and Technology Services Department dedicated support to that effort with technology integration. The literacy program emphasized professional development for writing instruction using models like the Calkin’s (2003) Writer’s Workshop. Inspired Writing became the initiative to support writing instruction through the use of inexpensive netbook computers assigned to Language Arts classrooms in all schools across the District.

During the initial pilot year in 2007-2008, five elementary schools deployed netbook computer carts in fifth grade classrooms in support of the new writing initiative. An independent grant evaluator concluded that achievement gains had been realized on three different forms of student assessment (Nebelsick-Gullett, n.d.). The district expanded the Inspired Writing initiative in 2008-2009 to every fifth grade classroom and every 6th and 9th grade Language Arts section in the district. An independent review of achievement data concluded that achievement differences between grade levels that were included in the Inspired Writing initiative saw achievement gains as high as 14% on the state writing assessments. The review concluded that the combination of the curriculum, professional development, and infusion of technology produced the achievement results (Warschauer, 2011).
The current state of the Inspired Writing initiative in Littleton Public Schools has been expanded to serve all students grades 5 through 12. LPS has purchased 2,800 netbooks to support the Inspired Writing program. In addition, school-initiated efforts to expand netbook availability in the classroom have resulted in an additional 5,700 computers district-wide. The total inventory of 8,500 Linux-based netbooks compliments the 5,500 Microsoft Windows-based computers maintained in the district. The district manages a domain for Google Apps for Education in which students across the district produced over 200,000 Google Docs during the 2011-2012 school year and over 180,000 documents during the fall semester of 2012-2013. The methodology of the initiative was documented and compared to other national examples as a model program in Warschauer’s (2011) Learning in the Cloud.

The one-to-one mac school district

The school district utilizing a traditional one-to-one laptop program reports 39% of its students qualify for the Free and Reduced Lunch Program and its minority students total 48% of the total student population. This school district entered into a contractual agreement with Apple Computers to provide all of its high school students with laptops. The program was rolled out incrementally, beginning with just the incoming freshman during the 2008-2009 school year and adding a class each year. Due to the fact that Colorado is a school of choice state, the principle purpose behind the infusion program in this school district was to recruit students from neighboring school districts. This school district is relatively small and the infusion effort has resulted in an annual increase in student enrollment by roughly 5%.

The non-infusion school district

The non-infusion school district is one of Colorado’s largest school district with an enrollment of more than 28,000 students. There are over 50 school sites. The district has a diverse population with the largest ethnic groups being 66% White and 28% Hispanic. Student access to technology varies greatly from school to school in this school district. Funding for technology has been provided by the district technology department and from parent/teacher organizations. The majority of technology funding has gone into infrastructure such as projectors, document cameras, and computer labs in schools. As of the 2012-2013 school year, there was not a plan to provide additional student access to laptops or netbooks in a unified manner.

Research questions

Given the unique nature of the LPS laptop infusion program, the researchers posed three questions they wanted to answer to better understand this focused ubiquitous effort. The three research questions are listed below:

- How has the Littleton Public Schools (LPS) model for infusing technology into the classroom influenced learning?
- How has the LPS model influenced teaching?
- How cost effective is the LPS model?

Conclusions

There is a high degree of consensus related to the use of technology to enhance the learning experiences of students in the 21st Century. The question is how school districts can most effectively introduce a one-to-one computing program into classrooms to provide all students with a technologically enriched learning opportunity? The data collected in this study help to gauge the effectiveness of two different approaches at infusing this technology into the classroom.

Methodology

The three research questions each required a specific methodology to ensure the validity of the findings. However, the overall design, population, and data collection instruments are described first. Then the methodology used for each research question is described.
Design

Creswell (2012) stated, “you conduct a mixed methods study when you have both quantitative and qualitative data and both types of data, together, provide a better understanding of your research problem than either type by itself” (p. 535). The researchers selected the mixed methods design for this study to better document the impact the LPS laptop infusion program is having on instruction and learning. To rely exclusively on one type of data would have resulted in a less than thorough analysis of the LPS laptop infusion model.

Population

The population for most of this study consisted of the LPS teachers and students in grades 5 through 12 language arts classrooms. The data collected from this first population produced the answers to research questions one and two. The population for the third research question included three school districts.

Data collection instruments

There was a total of four data collection instruments used in this study. The first data source was interviews with LPS language arts teachers, grades 5 through 12. The second data source came from focus group discussions with LPS students enrolled in language arts classes, grades 5 through 12. The third data collection instrument was a quantitative classroom observation tool that objectively documented the number of pedagogical traits observed in language arts classrooms at specific time intervals. The final data collection instrument was the cost analysis formula.

Research question #1

For the first research question, teacher and student interview questions were created in order to evaluate the effects of technology on learning. In terms of teachers, the questions were focused on lesson objectives and the integration of technology as a tool for writing instruction. During the initial pre-conference interview with teachers, questions were intended to explore the amount of professional development they received related to netbook integration, their views on technology as a tool for learning, and whether or not teachers felt that netbook technology had altered the writing process for their students.

Student interviews were focused on the way students felt their writing had or had not changed as a result of the use of netbook computers in the classroom. Student interviews took place about half way through the school year after students had been given ample time to use netbooks during writing periods and were held in a focus group format during the school day. There was value in having students interact with one another as they related their collective experiences with the use of technology in the classroom environment. Both teacher and student interviews were conducted in a semi-structured manner with key questions and sub-questions based on participant responses.

Research question #2

In order to answer the second research question, an observation tool was created with the intent of recording interactions that took place in a one-to-one computing rich classroom. The observation tool was based on partial interval recording with data being collected every five minutes during a 40-minute period. Individual indicators were checked if 80% of the class exhibited the behavior during each five-minute interval.

On the observation tool, the indicators are separated into three areas: (1) Use of Technology, (2) Content, and (3) Pedagogy. The Use of Technology section has six observable indicators, with eight periods of observation, for a total score of 48. The Content section has three observable indicators, with eight periods of observation, for a total score of 24. Finally, the Pedagogy section has four observable indicators, with eight periods of observation, for a total score of 32 (the observation tool is available upon request).
The standard utilized from the North Carolina teacher evaluation tool, labeled *Teachers Facilitate Learning for Their Students*, which served as a second data point, has eight indicators based on a four-point scale for a total overall score of 32 (McREL, 2009). Data were collected for this tool during the 40-minute observation period and scored when a behavior was observed. Evidence supporting each of the eight indicators was recorded based on student/teacher interactions on a separate sheet as a way to collect additional data resulting in final scores on the North Carolina scale.

After each classroom observation, the total score for the three sections on the observation tool were compared to the overall score on the North Carolina scale. In order to analyze the data, a bivariate correlation was conducted. This analysis determined whether or not the three independent variable, 1) effective use of technology, 2) strong writing content, and 3) sound pedagogy, had a positive or negative relationship with the dependent variable identified on the North Carolina scale, that of teacher effectiveness in the classroom setting. The data was also triangulated for trustworthiness through the use of observations, interview and focus groups, and by conducting member checks with the interview data collected.

**Research question #3**

A formula was developed to quantify the cost effectiveness of the different laptop infusion programs detailed in this study. The formula had to control for the size of the school district since the three districts in this study vary in student population. The formula divided the cost of the infusion program by the number of students served by the infusion program to generate a per pupil expenditure figure:

\[
\text{Cost of Infusion Program} \div \text{Students Served} = \text{Infusion Program Per Pupil Expenditure}. 
\]

The formula was applied to the data from all three districts for the last three years to measure the trends related to the per pupil expenditure of each infusion program.

**Findings**

The findings presented in this section are structured around the three research questions. These findings will determine the influence the LPS one-to-one computing program had on learning and teaching. In addition, the cost analysis findings are presented to measure the fiscal impact of each approach to infusing computers into the classroom.

**Findings: Student learning**

Based on the teacher and student interviews, certain themes emerged largely in favor of the use of netbooks for writing. An analysis of the data encapsulated through the coding process identified five themes. The first theme was focused on technology supporting a student’s ability to revise and edit their work. Teachers talked about the ease of revision specific to the writing process, which is often a roadblock for many reluctant writers. Students discussed how the use of netbooks allowed them to change their work with very little effort. For many teachers, the efficiencies related to the editing process also appeared to free students up to write more and to take risks with their writing that they may not have done with a handwritten product.

The second theme highlighted the importance of professional development for teachers and the need for quick and effective support. In this case, teachers highlighted the extensive training they received regarding the use of netbooks for writing instruction. LPS had a unique approach to professional development in that they provided a lot of guidance to teachers initially through required trainings and then provided ongoing support to teachers throughout the school year. In addition to their initial training, teachers were given opportunities to participate in classes and to share ideas with one another during the school year. This open dialogue helped teachers to take risks and make mistakes, which are key components of the learning process. Teachers also discussed the value in having technology support within their school building. Quick support appeared to minimize teacher frustration and prevent teachers from giving up on the netbooks altogether.
Differentiation was another key component that surfaced through the teacher and student interviews. Students talked about accessing information quickly in a way that enhanced the learning process. A key finding in this area related to students becoming less dependent on the teacher and more self-directed with their learning. In fact, some students made direct comments about how they were beginning to rely less on their teacher to complete their work. One student stated: “Like if you forgot the directions, you can just look back at them (in Google Docs), like the kids don’t have to go ask the teachers for the directions again.” Students who were less dependent on their teacher also talked about being more engaged in the learning process. One student talked about engagement this way:

I’m more interested because I’m a lot better at computers than just watching a teacher because watching the teacher for like 15 and 20 minutes makes me really bored because she’s just going “blah, blah, blah, blah, blah”...and when we do our netbooks, we’re not just watching our teacher...but we’re actually doing stuff.

While engagement is a difficult concept to quantify, self-directed learning, according to the students who were interviewed, did in fact help them be more engaged in what was going on in class.

Teachers mirrored student comments in terms of the way netbook use helped them to differentiate instruction for their students. During some of the interviews, teachers talked about work with students, which allowed them to guide individuals through the learning process rather than just being in front of the class teaching. According to one teacher, “I mean if you’ve bought a kid a new toy…and for two weeks it’s the greatest thing on earth and then nothing, you know. And it would be the same thing with the netbooks if you just dropped them into a room. It has to change the way you teach and that’s what makes the engagement.” In these instances, the teacher’s role had changed from “sage on the stage,” as one teacher put it, to a true facilitator of learning. This finding was supported through the observation process as well. Those teachers who were rated as being most effective on the observation tool spent little time teaching in a traditional manner. Instead, they provided quick instructions to the class, which allowed students to access information pertinent to the lesson on their netbooks. The teachers then moved around the room to support students individually, which is a key aspect of differentiated instruction. Similarly, an analysis of the interview data indicated that teachers needed to give up some control in the traditional teaching and learning process. Teachers who were willing to take risks and struggle right along with their students experienced success and even looked to students to help guide instruction. Empowering students appears to lead to increased student engagement throughout the learning process.

Another prominent theme that emerged was the increased ability for teacher to student and peer-to-peer feedback in the area of writing. By posting student writing in an electronic manner, both teachers and students were able to view written work and give comments to one another to make their writing better. For teachers, this type of feedback allowed them to interact with students more frequently and freed them up to make multiple comments that a student could view and follow up on later. This was in direct contrast to a traditional writing conference where students might confer with a teacher and then go back to their desk to write. The main points of the conference in this more traditional format are often lost or forgotten. Students also appeared to be very comfortable looking at each other’s work when netbooks were in use and providing meaningful feedback on what was written. This additional feedback meant that several people were looking at what had been produced in addition to the classroom teacher.

Finally, technology provided an expanded audience for student writing. Teachers talked about sharing student writing electronically with parents, grandparents and other adults. This expanded audience helped students to realize that what they were writing was public and was being viewed by many individuals at any given time. Teachers commented on how this expanded audience helped students to put more effort into what they were writing. Wikis and blogs were used in many of the classrooms as a format for students to share their work with others.

Findings: Teaching

In addition to the teacher and student interviews, data collected from the use of the observation tool helped to answer the research question focused on student engagement when technology was in use in classrooms as a tool for learning. The findings from the 18 classroom observations indicated that students were more engaged in the learning process when teachers utilized technology in a successful manner. That is, teachers who scored the highest on the North Carolina scale, intended to gauge a teacher’s ability to facilitate learning in the classroom, tended to use netbook technology to teach writing in a more effective manner than those teachers who scored lower on the North
Carolina scale. This finding suggests that the need for integration of pedagogy and technology is valuable in enhancing student learning.

Effective and efficient use of netbooks as a tool for writing also resulted in increased student engagement as measured through higher scores on the observation tool in terms of a teacher’s ability to use technology while also providing strong content and pedagogy for students. Furthermore, netbook use for the highest rated teachers resulted in an enhanced written product. To put it simply, student work was enhanced by the effective use of technology use in the classroom. This is an important finding and suggests that students can become better writers when netbooks are in use and that technology, when used effectively, may be more than just another glorified tool for writing in schools.

Findings: Cost effectiveness

Despite the fact that the LPS infusion model positively impacts student achievement and teaching, the focus of the third research question is on the cost effectiveness of the different approaches to infusing laptops into the classroom. Table 1 summarizes the per pupil expenditures for each of the three school districts included in this study.

<table>
<thead>
<tr>
<th>School Year</th>
<th>Cost</th>
<th>Students Served</th>
<th>PPE Infusion</th>
<th>Cost</th>
<th>Students Served</th>
<th>PPE Infusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2010</td>
<td>$282,000</td>
<td>2,785</td>
<td>$101.26</td>
<td>$155,897</td>
<td>189</td>
<td>$824.85</td>
</tr>
<tr>
<td>2010-2011</td>
<td>$274,000</td>
<td>2,963</td>
<td>$92.47</td>
<td>$151,837</td>
<td>375</td>
<td>$404.90</td>
</tr>
<tr>
<td>2011-2012</td>
<td>$315,000</td>
<td>4,437</td>
<td>$77.99</td>
<td>$142,818</td>
<td>548</td>
<td>$260.62</td>
</tr>
</tbody>
</table>

The cost of the two computer laptop infusion efforts decreased as more students benefitted from the efforts. However, using the PPE Infusion formula, which divided the cost of the infusion effort by the number of students being served, it is clear to see that the LPS model is more cost effective than the more traditional infusion effort. The cost effectiveness of the LPS model is accentuated by the fact that its most expensive PPE Infusion, $101.26 per student, was still only 39% of the lowest PPE Infusion cost for the traditional infusion model.

Implications

There are several implications that can be drawn from the analysis of these data and assimilated for teachers, students, and principals. First and foremost, classroom teachers need to understand that they can, in fact, have a positive impact on the way students write when using netbook technology in the classroom setting. The results of this study indicate that effective teachers can use netbooks to differentiate instruction for students and expand the scope of individuals providing feedback to students about their writing. The data also suggest that teachers should take risks and experiment with the various tools available to them including wikis, blogs, and Google Docs to provide templates and guidelines for students that allow them to become more self-directed and responsible for their learning. In addition, teachers who embrace the use of netbooks for writing need to re-examine how they teach. The data indicate that the most effective teachers in this study have begun to shift away from the traditional teaching model – that of presenting information to a group of students from the front of the class with limited interaction or opportunities for feedback. Instead, these teachers are presenting material in a way that requires very little direct instruction so that students know where to go and what to do, which in turn frees up a teacher to conference individually with more students or even coach students on how to provide meaningful feedback to one another.

The students who took part in this study appeared to have increased confidence as writers when netbooks were in use. They talked about being more self-directed and also about the value of interacting with a variety of individuals,
including their classroom teacher, throughout the writing process. One student said, “I think that the netbooks are good because it’s easier to get feedback from teachers and from one another on our blogs that we make for reading.” Another student said, “It definitely helps because if you want to check someone else’s work, all they have to do is share it with you so that you can make sure that their work is good just like yours is.” Obviously these students had the benefit of hands on instruction as to the proper and effective use of netbook technology, but it was impressive to witness firsthand how competent 5th grade students had become after only a few months of using netbooks in the classroom. Day-to-day access to this technology in conjunction with a teacher dedicated to netbook use appeared to be a key aspect of some students’ success as well, which is an argument in favor of the one-to-one model. For the classrooms that were observed, the transition from a reading or math lesson to a writing lesson contingent on netbook use was seamless and part of the daily routine for students.

Another important implication is that students, much like teachers, need to have the opportunities to make mistakes and take risks when technology is in use. During classroom observations, a number of students were seen interacting with one another through electronic means, asking for help, or attempting to try out a new procedure or process. The teachers in this study stressed that students learned best through trial and error and when given the opportunity to experiment they frequently experienced growth as writers.

Finally there are various implications for principals and other school leaders when consideration is given to netbook use in schools. A key point involves the scope of professional development provided to teachers. In this study, teachers talked a lot about the benefit of ongoing professional development as well as the opportunity for quick and efficient technical support. Teachers who received training focused on the use of netbooks and the use of technology tools related to writing appeared to have a positive attitude about the various ways this type of technology could be used in the classroom. This is in direct contrast to some one-to-one programs where teachers are given a netbook or laptop with very little instruction or support and basically encouraged to figure it out on their own.

School principals and leaders should also take the time to obtain feedback from teachers on how netbook technology is working as a tool for learning. Principals simply cannot assume that technology will make a positive impact on the way students learn. If a principal knows how netbooks are being used, effectively or ineffectively, support can be provided to increase a teacher’s capacity to positively impact student performance in writing and in other subject areas as well.

Conclusions

Based on the findings detailed in this study, the LPS Inspired Writing infusion effort is a model that other school districts should replicate. The data reported here demonstrate that the Inspired Writing program positively impacts writing and teaching. In addition, the LPS model has been shown to be more cost effective than providing the more traditional ubiquitous laptop programs. The title of this study purposefully includes the phrase “focused ubiquity.” The LPS laptop infusion program is commendable for its focused approach to positively influence student achievement. Instead of just providing students with laptops and all teachers with some professional development, LPS educational leaders decided to focus the potential impact of laptops in the classroom on improving writing skills and, as a result, focused all professional development on how language arts teachers could best utilize this technology. The ubiquitous nature of the LPS model lies in the fact that the laptops are made available to all students in their language arts classes. Realizing that all school districts operate on a limited budget, any infusion effort that produces the results reported in this study at a fraction of the cost of a traditional laptop infusion effort is worthy of emulation.

References


Enhancing Students’ NOS Views and Science Knowledge Using Facebook-based Scientific News

Hsi-Yu Huang, Hui-Ling Wu, Hsiao-Ching She* and Yu-Ren Lin

Institute of Education, National Chiao-Tung University, Taiwan, No. 300, University Road, Hsinchu, Taiwan 1001 // yangamy0915@yahoo.com.tw // kellyhwu@gmail.com // hshe@mail.nctu.edu.tw // titiyoyo@msn.com

*Corresponding author

(Submitted July 19, 2013; Revised October 28, 2013; Accepted January 27, 2014)

ABSTRACT

This study investigated how the different discussion approaches in Facebook influenced students’ scientific knowledge acquisition and the nature of science (NOS) views. Two eighth- and two ninth-grade classes in a Taiwanese junior high school participated in the study. In two of the classes students engaged in synchronous discussion, and in the other two classes they engaged in asynchronous discussion. The 83 participating students discussed scientific news that had been posted on Facebook, with a total of seven news stories. Three instruments – a general NOS views questionnaire, a content-related NOS views questionnaire, and a science knowledge test – were administered before, immediately after, and three weeks after the intervention. In addition, student responses to each news story were collected and analyzed. The results demonstrated that regardless of which online discussion group they participated in, students improved their science content knowledge, general NOS views, and NOS analyses of scientific news. Especially, synchronous discussion proved more effective than asynchronous discussion at helping students to develop science content knowledge and increase their view of community aspects of content-related NOS. These findings demonstrated the advantages of quick feedback and responses in synchronous discussion facilitated communication and cognitive development about subject and subject-related nature of science.

Keywords

Nature of science views, Science knowledge, Facebook, Scientific news

Introduction

Studies have consistently found that students usually lack an adequate understanding of the nature of science (NOS) (Lederman, 1992; Ryan & Aikenhead, 1992). The nature of science is defined as the epistemology of science, characteristics of scientific knowledge, and the values and beliefs inherent to scientific knowledge and its development (Lederman, 1992). Many of these intrinsic ideas are not emphasized in traditional science classroom instruction, resulting in students learning skewed notions about how science is conducted (Abd-El-Khalick & Lederman, 2000). Exploring ways to help students acquire an adequate understanding of NOS is important for educators, since this is among the desired outcomes of science instruction (Lederman, 1992; Abd-El-Khalick & Lederman, 2000). The mass media tend to cover contemporary socio-scientific issues which address the public’s attention and interests as they are related to daily life (Halkia & Mantzouridis, 2005). The values of media news stories in science instruction seem promising given that the use of scientific news stories in education prepares individuals to learn about science and to build essential knowledge of NOS (Christensen, 2011; McClune and Jarman, 2012). In sum, NOS and news stories about science are interrelated and have an effect on one another. One particular media form has gained prominence recently as a means of conveying relevant information throughout society: social networking. Popular social networking sites such as Facebook can facilitate collaborative science learning because they are user-friendly and support flexible communication (Baatarjav, Phithakkitnukoon, & Dantu, 2008), thus providing ideal spaces for users to engage intellectually and form networked learning communities. Although researchers generally agree that explicit instruction improved students’ NOS learning, few have examined how different online communication formats influence students’ understanding of NOS. Knowledge remains limited regarding how reading and discussing scientific news in online contexts affects students’ NOS comprehension. This study compared how students’ NOS views changed when they analyzed scientific news stories in synchronous and asynchronous online discussions. The findings of the current study reiterate the value of scientific news stories in informal science education and broaden our knowledge about how news media influences individuals’ awareness of NOS.
Literature review

The rising influence of Facebook in instruction

Social networking sites, such as Facebook and MySpace, are currently among the most popular online communication methods. The number of Facebook users has grown exponentially from 1 million in 2004 to over 1 billion in 2012 (Smith, Segall, & Cowley, 2012), compelling some researchers to explore the potential educational benefits of using such sites. One advantage of using Facebook was that it offered learners “a dynamic and unintimidating environment” in which to communicate (Schroeder & Greenbowe, 2009, p. 5), and social networking has the potential to support diverse instructional strategies, dynamic learning materials, and learning communities (Arnold & Paulus, 2010). Scholars thus have new opportunities to explore how social networking sites can be used to influence instructional design and learning outcomes.

Synchronous and asynchronous online discussion

Online discussion has been shown to benefit student learning by fostering high-level cognitive skills such as reasoning, argumentation (Pilkington & Walker, 2003; Yeh & She, 2010), and subject comprehension (Comeaux & McKenna-Byington, 2003; Chen & She, 2012). Researchers continue to explore the educational advantages and disadvantages of different online communication approaches, particularly synchronous and asynchronous discussion.

Synchronous discussion involves online interaction in real time, establishing urgent and immediate feelings among students that foster a sense of community and that works well for content that stimulates debate (Schwier & Balbar, 2002). Learners enjoyed synchronous online discussions because it provided more social interaction (Hrastinski, 2008). It also fostered both cognitive and social activities (Oztok, Zingaro, Brett, & Hewitt, 2013). Despite receiving quick feedback (Davidson-Shivers, Muilenburg & Tanner, 2001), students had difficulty keeping up with long threads of discussion (Khine, Yeap, & Lok, 2003) and engaging in in-depth thinking (Hrastinski, 2008) when using synchronous online discussion within a short period of time.

While synchronous online discussion promotes the “social” aspect of education, asynchronous communication supports “academic” aspects of learning (Motteram, 2001). It is more appropriate for learners to engage in asynchronous online discussions when engaged with complex ideas because they have sufficient time to respond to posted messages (Hrastinski, 2008; Mandernach, Dailey-Hebert, & Donnelli-Sallee, 2007, Schrire, 2006).

The nature of science

The nature of science (NOS) is an all-encompassing concept that includes the process of forming scientific knowledge, and concerns “how scientists operate as a social group and how society itself both directs and reacts to scientific endeavors” (McComas, 2008, p. 250). It contains one’s “conceptions about scientific knowledge and knowing, values and beliefs incorporated in gaining scientific knowledge, as well as the influences of society, culture, and technology on science” (Urhahne, Kremer, & Mayer, 2011).

While scholars have no consensus concerning the aspects of NOS, they often note three important issues. First, science derives from the collection of accumulated knowledge. New scientific claims emerge when scholars reinterpret old evidence, explain existing theories differently, or engage in new research directions. As such, scientific knowledge never remains absolute and certain (Lederman, 2007). Second, science involves creativity and imagination. Scientists need much creativity to explain science (Lederman, 2007). One science education standard, Benchmarks for Scientific Literacy (AAAS, 1993), indicated that learners should apply their imaginations to devise hypotheses and explanations while attempting to make sense of collected evidence and construct evidence-based explanations. Finally, scientific endeavors are characterized by continuous discussion, communication, and criticism among scientists. Within a society, scholars often encounter situations in which their scientific claims are questioned, especially when they use inappropriate, insufficient, or misinterpreted evidence. Individuals must negotiate and communicate their ideas with others in order to reach consensus. When scientific communities have internal agreement, their conclusions are likely to be accepted by the public.
International science education organizations, such as the American Association for the Advancement of Science (1993) and the National Research Council (1996), recognize that an adequate understanding of NOS would help students in scientific literacy. In this regard, learner understandings of NOS continuously constitute an area of intense academic interest. However, studies show that science teachers and students do not always possess adequate conceptions of NOS (Halai & McNicholl, 2004; McComas, Clough, & Almazroa, 1998).

**Implicit and explicit NOS instruction**

The differences between implicit and explicit approaches lie in the extent to which learners think about and reflect on the specific aspects of NOS (Abd-El-Khalick & Lederman, 2000). Implicit instruction supports students to experience “what science is and how it works” but providing no direct content to understand NOS (Clough, 2006, p. 467). However, several studies argued that implicit NOS instructions were generally not successful (Sandoval & Morrison, 2003; Schwartz, Lederman, & Thompson, 2001).

Considering the limitations of implicit NOS instruction, researchers suggest teaching NOS with explicit strategies to promote student reflection and discussion. Explicit NOS instruction deliberately designs science lessons to draw student attention to important NOS issues without lecturing or imposing NOS perspectives (Clough, 2006). In addition, prior studies widely reported that explicit NOS instruction supported student learning better than implicit instruction. Creswell and Clark’s (2007) study also demonstrated explicit and reflective instruction that helped students to achieve better NOS views and content knowledge than implicit instruction produced. Similarly, Peters (2012) found that students in the explicit instructional group significantly outperformed those in the implicit group on NOS. Thus, this study employed explicit instruction for NOS instruction.

**Learning science through news stories**

Recognizing the limitations of formal science education curriculum tend to teach straightforward but isolated scientific facts, scholars assert the significance of using real-world sources to bridge the gap between school curriculum and science in everyday life. The experiences individuals have in various learning contexts, such as watching television, visiting museums, and reading newspapers and magazines, influence the ways “individuals construct scientific knowledge, attitudes, behaviors, and understanding” (Dierking, Falk, Rennie, Anderson, & Ellenbogen, 2003, p. 109). Thus, real-world experiences shape individual’s understanding of scientific knowledge and perceptions about the process of science. In addition, real-world learning experiences prepares students “to engage with science in the contexts they will encounter in later life.”(McClune & Jarman, 2010, p. 728). To achieve this, DeBoer (2000) suggests science education should prepare individuals to critically understand “reports and discussions of science that appear in the popular media” (p. 592).

For over a decade, instructors have recognized the potential value of utilizing media reports of scientific research in instruction to foster scientific literacy (Korpan, Bisanz, Bisanz & Henderson, 1997). According to McClune and Jarman (2010), almost any secondary school in the UK used newspapers as a resource to support science instruction to some extent. Given the fact that public media usually reports socio-scientific issues that are related to daily life, using newspapers in science education not only communicates important science information with individuals but also keeps learners informed about cutting-edge science issues. In this respect, news stories offer a great opportunity to stimulate learning interests, which facilitates further research and lifelong learning (McClune & Jarman, 2012). Especially, scholars (Korpan, Bisanz, Bisanz & Henderson, 1997; Zimmerman, Bisanz, Bisanz, Klein, & Klein, 2001) emphasized that media-based scientific research reports serve as an important source of new science knowledge in an age of information. Korpan and colleague’s (1997) study demonstrated the benefits of science-based news stories, finding that through reading, evaluation, and discussion about scientific research and news, readers not only learned the news content but also realized what the scientific community worked for and how scientific knowledge is formed.

**Purposes of the study**

The present study aims to investigate how synchronous and asynchronous online discussions via Facebook influence student informal science learning, especially on content-related NOS and science content knowledge. Three research questions guided the present study:
• Do synchronous and asynchronous discussions of scientific news stories influence student learning differently, particularly as regards the students’ general NOS views, content-related NOS views, and science content knowledge?

• How do different online discussion approaches (synchronous and asynchronous) influence students’ accumulative, creative, and community aspects in NOS when they analyzed scientific news stories?

• How do general NOS views and science content knowledge contribute to the development of content-related NOS views?

Methods

Participants

The current study took place in a junior high school in Northwestern Taiwan. Two 8th-grade and two 9th-grade science classes, all of which were taught by the same instructor, participated in this study. In total, 49 male and 34 female students joined the online discussion activities. The t-test gained from three science examinations in the first semester of the 8th grade year indicated that the participating students in both synchronous (N = 42) and asynchronous (N = 41) groups had no significant difference in their prior knowledge (t(83) = 0.18, p = 0.86).

Procedure

The current study encompassed eight class periods lasting 45 minutes each and occurring over the course of three weeks. Following the completion of three pretests (general NOS views questionnaire, content-related NOS views questionnaire, and science knowledge test), the instructor explained to the students the content of learning activities as well as features of NOS and the significance of understanding it. To help students understand NOS, this study posted seven scientific news stories on Facebook which allowed students to analyze the accumulative, creative, and community of NOS in each designated scientific news story (see Figure 1). The stories consisted of two themes: heat (calorie, exercise, green house) and elements (environmental change, gold, proton, high temperature superconducting material). The day after the students finished the online learning activities, they were given three multiple-choice posttests (general NOS views questionnaire, content-related NOS views questionnaire, and science knowledge test). In addition, in order to test whether students retained their knowledge, the same three tests were administered again three weeks later.

Figure 1. Scientific news story on Facebook.
Research design

This study adopted a quasi-experimental design which randomly assigned the four science classes to synchronous or asynchronous groups. One class in each grade level engaged in synchronous discussion, and the other engaged in asynchronous discussion. Each class was divided into five groups, with each group consisting of four to five students. While students in the synchronous groups spent 45 minutes (one class) analyzing a scientific news story on Facebook (see Figure 2), students in the asynchronous group had three days to finish their discussion of one news story.

Figure 2. Students’ discussion on Facebook: Scientific news story.

The participating students logged-in to Facebook and navigated to the page devoted to this study. For each discussion topic, the instructor posted the text of a scientific news story and a video clip related to the story. Following these, students were provided with three questions that required them to analyze the NOS features (accumulated knowledge, creativity, and scientific community) in each story. For example, students were asked “Which descriptions explained the creativity of science knowledge?” in the news story. In addition, each news story included one question encouraging students to discuss the content of the story, such as “Why isn’t it appropriate for patients with kidney problems to drink tea?”

Data collection

General NOS views questionnaire

Based on the important features of NOS identified in prior studies (Chen, 2006; Kao, 2006; Tsai, 2007), a questionnaire was designed to explore student general views of three features of NOS: accumulative, creative, and community. Fifteen content-independent multiple choice items, five for each NOS feature, investigated general NOS views. One of the items was, “A new scientific theory must be recognized by the majority of scientists in order to achieve the validity.” The instrument was measured by a 5-point Likert scale, ranging from strongly agree (5 points) to strongly disagree (1 point). The validity of the questionnaire was established by the input of four science education doctoral students and two scholars. The overall reliability of the questionnaire items was 0.86. Each of the three NOS features were represented on the questionnaire by five items each, with satisfied reliability (accumulative = 0.83, creative = 0.69, community = 0.82).
**Content knowledge test**

The content knowledge test consisted of fifteen multiple choice questions. The test measured students’ comprehension of the science content they discussed in the online interactions. This test required students to understand each discussion topic (2 items for each) and the practice topic before the intervention (1 item). In addition to the established expert validity, a Cronbach’s alpha showed adequate reliability scores for the comprehension measure (0.76).

**Content-related NOS views**

Two measurement methods were applied in the current study for analyzing students’ understanding of content-related NOS while analyzing scientific news stories. First, this study implemented a newly-designed instrument (Content-related Nature of Science View Questionnaire) that was based on scientific news stories. These stories were discussed in class, and were used to assess the changes in students’ content-related NOS views. Unlike the general NOS views questionnaire, this one measured how students analyzed the NOS in each scientific news story, specifically regarding the features of accumulated knowledge, creativity, and scientific community. Within each scientific news story, the three major NOS features were measured individually using two multiple-choice questions. First, students justified which description in the selected paragraph best explained the accumulated knowledge, creativity, or community aspects of NOS. Second, students explained their reasons by answering the multiple-choice question. For instance, a question asked,

*Why do you think your choice best describes the feature of NOS? Please explain.* (A) They designed an unprecedented biological experiment because they used monkeys as subjects. (B) The experiment lasted for more than twenty years. (C) The cause of the monkeys having different appearances was their diet. (D) The research team pointed out an innovative theory: longevity comes from a low calorie diet.

Students received full credit (one point) only if they answered both questions accurately. If not, they gained no points. The questionnaire contained 42 questions with an average reliability of 0.83.

Second, the discussions in the Facebook forum were measured to evaluate the level of student understanding about NOS in each aspect. During the discussion, students analyzed each scientific news story by identifying its NOS features in terms of three major aspects of accumulated knowledge, creativity, and community. A science educator and a middle school science teacher developed a coding system to analyze student NOS views for each scientific news story. Two middle school science teachers used this coding system to categorize students’ responses into three levels for each aspect of accumulated knowledge, creativity, and community – complete (2 points), partially complete (1 point), and unrelated (0 point) – with an inter-rater reliability of 0.93.

**Results**

This study investigated how different discussion approaches in Facebook influenced students’ scientific knowledge acquisition and the nature of science (NOS) views both in general and content-related views. A multivariate analysis of the questionnaires given before the intervention showed that the performance of students in both the synchronous and asynchronous groups showed no statistically significant difference in terms of content knowledge, and general and content related NOS views ($F(3, 79) = 0.65, p = 0.59, \text{Wilk's } \lambda = 0.98, \text{partial } \eta^2 = 0.02$).

**General NOS views**

The pretest, posttest, and retention test scores were collected to measure whether students’ general NOS views have changed significantly after they engaged in synchronous or asynchronous online discussions. A paired samples $t$ test in Table 1 indicated that after participating in online discussions, students on average achieved significantly better understanding of NOS as indicated by the different posttest and pretest scores, regardless of whether they were assigned to a synchronous ($t (41) = 2.14, p = 0.04, \text{Cohen’s } d = 0.33$) or asynchronous ($t (40) = 2.12, p = 0.04$, 294
Cohen’s $d = 0.33$) group. However, only synchronous online discussion groups showed a significant difference between the pretest and retention test scores ($t (41) = 2.88$, $p = 0.01$, Cohen’s $d = 0.44$).

Table 1. Descriptive statistics and paired $t$-test

<table>
<thead>
<tr>
<th></th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Retention M(SD)</th>
<th>Posttest- Pretest $t$ test($p$)</th>
<th>Retention- Pretest $t$ test ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Views of NOS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchronous (N = 42)</td>
<td>56.55 (6.45)</td>
<td>59.05 (5.99)</td>
<td>59.33 (6.25)</td>
<td>2.14 (0.04)</td>
<td>2.88 (0.01)</td>
</tr>
<tr>
<td>Asynchronous (N = 41)</td>
<td>56.51 (6.34)</td>
<td>59.41 (7.23)</td>
<td>57.49 (6.71)</td>
<td>2.12 (0.04)</td>
<td>0.78 (0.44)</td>
</tr>
<tr>
<td><strong>Science knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchronous (N = 42)</td>
<td>4.52 (2.64)</td>
<td>10.05 (2.52)</td>
<td>10.17 (2.77)</td>
<td>12.24 (0.00)</td>
<td>11.22 (0.00)</td>
</tr>
<tr>
<td>Asynchronous (N = 41)</td>
<td>3.88 (1.81)</td>
<td>8.39 (3.81)</td>
<td>8.27 (3.46)</td>
<td>7.88 (0.00)</td>
<td>8.19 (0.00)</td>
</tr>
<tr>
<td><strong>Content-dependent NOS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchronous (N = 42)</td>
<td>2.10 (1.22)</td>
<td>7.95 (4.59)</td>
<td>7.83 (5.45)</td>
<td>8.81 (0.00)</td>
<td>7.67 (0.00)</td>
</tr>
<tr>
<td>Asynchronous (N = 41)</td>
<td>1.85 (1.74)</td>
<td>7.49 (4.92)</td>
<td>7.05 (5.16)</td>
<td>7.58 (0.00)</td>
<td>6.45 (0.00)</td>
</tr>
</tbody>
</table>


Multivariate analyses of covariance (MANCOVA) were conducted to investigate whether instructional strategies (synchronous vs. asynchronous discussion) influenced students’ overall general NOS views on both the posttest and retention test, with the pretest as the covariate in each analysis model. Instructional strategies were entered as the independent variable and the general NOS views questionnaire as the dependent variable. The preliminary data screening indicated no violation of the assumptions of linearity, homogeneity of variance-covariance matrices, and equality of variance. The result showed that there was no statistically significant difference between synchronous and asynchronous discussion groups on the combined dependent variables after adjusting the pretest score, $F(2, 79) = 1.98$, $p = 0.15$, Wilk’s $\lambda = 0.95$, partial $\eta^2 = 0.05$ (see Table 2). In short, students in the two discussion groups did not have significantly different general NOS views in either the posttest or retention test, after adjusting the pretest scores.

Table 2. Multivariate analysis of covariance in views of NOS, content-related NOS view, and science knowledge tests

<table>
<thead>
<tr>
<th></th>
<th>Wilk’s $\Lambda$</th>
<th>$A$</th>
<th>Multivariate $F$ ($p$)</th>
<th>Univariate $F$ Posttest ($p$)</th>
<th>Univariate $F$ Retention ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Views of NOS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>0.85 (0.15)</td>
<td>6.80 (0.00)</td>
<td>0.07 (0.79)</td>
<td>1.92 (0.17)</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.95 (0.05)</td>
<td>1.98 (0.15)</td>
<td>0.91 (0.09)</td>
<td>3.84 (0.03)</td>
<td></td>
</tr>
<tr>
<td><strong>Science knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>0.91 (0.09)</td>
<td>3.84 (0.03)</td>
<td>0.93 (0.07)</td>
<td>3.13 (0.05)</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.93 (0.07)</td>
<td>3.13 (0.05)</td>
<td>0.93 (0.07)</td>
<td>3.13 (0.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Content-related NOS views</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>0.89 (0.12)</td>
<td>5.11 (0.01)</td>
<td>0.89 (0.12)</td>
<td>5.11 (0.01)</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>1.00 (0.00)</td>
<td>0.17 (0.84)</td>
<td>1.00 (0.00)</td>
<td>0.17 (0.84)</td>
<td></td>
</tr>
</tbody>
</table>

**Content knowledge**

As the pretest/posttest scores in Table 1 show, students improved their science content knowledge significantly after participating in this study, regardless of whether engaged in synchronous ($t (41) = 12.24$, $p = 0.00$) or asynchronous
discussion ($t(40) = 7.88, p = 0.00$). Comparing the pretest and the retention test showed significant student knowledge growth in both the synchronous ($t(41) = 11.22, p = 0.00$) and asynchronous groups ($t(40) = 8.19, p = 0.00$).

A MANCOVA analysis tested whether students in different discussion groups achieved significantly different levels of science knowledge comprehension in the posttest and retention test when students’ prior test was controlled as the covariate. The assumption of the homogeneity of variance was met. The analysis revealed that instructional strategies had a significant effect on the combined dependent variables, $F(2, 79) = 3.13, p = 0.049$; Wilk’s $\lambda = 0.93$; partial $\eta^2 = 0.07$. Follow-up univariate ANCOVAs continued to showed that instructional strategies in online discussion groups significantly affected students’ science comprehension immediately after learning (Welch’s $F(1, 69.15) = 5.45, p = 0.02$). Students in the synchronous discussion group achieved higher content knowledge than those in the asynchronous discussion group in the posttest. The positive effects of synchronous discussion over asynchronous discussion on science knowledge development were also reflected in the retention science content knowledge test, Welch’s $F(1, 76.49) = 7.61, p = 0.01$.

### Content related NOS views

#### Content-related NOS Views Questionnaire

Paired sample $t$ tests were conducted to evaluate the impact of online discussion on student scores on the Content-related NOS Views Questionnaire. There was a statistically significant increase between the pretest and posttest questionnaire scores in both the synchronous ($t(41) = 8.81, p = 0.00$, Cohen’s $d = 1.36$) and asynchronous ($t(40) = 7.58, p = 0.00$, Cohen’s $d = 1.18$) groups. Comparing pretest and retention test scores also indicated that student overall content-related NOS views improved after they participated in this study, regardless of to which group they were assigned ($p = 0.00$).

A MANCOVA was performed to investigate whether different instructional strategies utilized in online discussion groups influenced students’ content-related NOS views. The MANCOVA test revealed that there was no statistically significant difference between synchronous and asynchronous discussion groups on the combined dependent variables, controlling for the pretest ($F(2, 79) = 0.17, p = 0.84$; Wilk’s $\lambda = 1.00$; partial $\eta^2 = 0.00$).

### Scientific news stories analysis in Facebook

The student discussions in Facebook were collected and analyzed to determine the extent of student understanding about each aspect of content-related NOS views. To determine whether online discussion strategies influenced student understanding of each aspect, three separate one-way ANOVA were conducted. The dependent variable in each statistical analysis model was student responses to scientific news stories in Facebook in terms of each aspect of NOS (accumulative, creative, or community aspects). Statistics found that different online discussion approaches influenced student comprehension of scientific community significantly, $F(1, 81) = 8.64, p = 0.00$ (see Table 3). When students participated in synchronous online discussion, they were more likely to gain a better understanding of scientific community concepts in content related NOS than those in asynchronous online discussion groups ($M_{\text{Synchronous}} = 1.19, SD = .08; M_{\text{Asynchronous}} = .88, SD = .08$).

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Mean (SD) Synchronous</th>
<th>Mean (SD) Asynchronous</th>
<th>$F$</th>
<th>$P$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulation</td>
<td>0.64 (0.06)</td>
<td>0.68 (0.06)</td>
<td>0.24</td>
<td>0.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Creativity</td>
<td>1.11 (0.72)</td>
<td>1.10 (0.07)</td>
<td>0.01</td>
<td>0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>Community</td>
<td>1.19 (0.75)</td>
<td>0.88 (0.08)</td>
<td>8.64</td>
<td>0.00</td>
<td>0.10</td>
</tr>
</tbody>
</table>

To understanding better the differences between synchronous and asynchronous discussion groups vis-à-vis the community aspect of content-related NOS, student responses to each scientific news story in Facebook were analyzed qualitatively.
At the beginning of the intervention, students in both groups tended to simply define scientific community as a teamwork system which encouraged scientists to discuss and conduct research cooperatively. Students in the asynchronous group continued this belief by defining scientific community as a group of people working together. For example, one student simply defined scientific community as "research teams." This indicated that students who engaged in asynchronous groups failed to define scientific community completely even after several discussions. In contrast, the majority of students in the synchronous group noted that one component of scientific community is the public. That is, other scientists had to comment on researchers’ ideas and the public must acknowledge these ideas and incorporate them into a public discourse. For example, student L described the community aspect in a scientific news story in the following way:

Professor Maw - Kuen Wu discovered a new research on the superconductor. His research had been recognized and accepted by other scientists and published to the famous journal in order to acquire the acknowledgement of scientific community. This is what scientific community means.

The benefits of a scientific community were also noted. Students in the synchronous groups stressed that researchers needed to record their studies carefully in order to allow scientists to examine and test their research outcomes.

Dr. Kennett’s research results were published in Science, the prestigious academic journal. Russian American scientist George Gamow re-examined Kennett’s theory with different scientific methods and reached similar conclusions. (Student H)

Thus, student view of community were close to what researchers (Lederman, 1992; McComas, 2008) emphasize; that is, that new knowledge had to be reported clearly and openly in order to establish community in NOS.

Predictions on content-related NOS views

In order to investigate how general NOS views and science content knowledge contributed to students’ development in content-related NOS views, three multiple linear regression models were carried out to determine the effects these variables had on the pretest, posttest, and retention test.

The first linear regression was performed to observe how pretests, which measured both science content knowledge and general NOS views, explained student content-related NOS views which were also measured before the intervention. Preliminary data checking showed that multicollinearity did not seem to be a problem for this model. The analysis found that the combination of independent variables significantly predicted students’ content-related NOS views, \( F(2, 80) = 5.12, p < .01 \). The R-square value indicated that about 12% of the variance in the dependent variable was explained by both independent variables. While the science knowledge pretest significantly predicted the students’ content-related NOS views as shown through their pretest scores (standardized beta = 0.35, \( p = 0.00 \)), NOS views alone did not appear to have a significant contribution to predicting students’ prior content-related NOS views (\( p = 0.27 \)).

<table>
<thead>
<tr>
<th>Table 4. Regression analysis results in views of NOS, content-related NOS view, and science knowledge tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regression 1: Pretest</strong></td>
</tr>
<tr>
<td>Outcome variable: Content-related NOS</td>
</tr>
<tr>
<td>Predictor variable: Science knowledge</td>
</tr>
<tr>
<td>( B )</td>
</tr>
<tr>
<td>0.31</td>
</tr>
<tr>
<td>0.04</td>
</tr>
<tr>
<td><strong>Regression 2: Posttest</strong></td>
</tr>
<tr>
<td>Outcome variable: Content-related NOS</td>
</tr>
<tr>
<td>Predictor variable: Science knowledge</td>
</tr>
<tr>
<td>( B )</td>
</tr>
<tr>
<td>0.78</td>
</tr>
<tr>
<td>0.13</td>
</tr>
<tr>
<td><strong>Regression 3: Retention</strong></td>
</tr>
<tr>
<td>Outcome variable: Content-related NOS</td>
</tr>
<tr>
<td>Predictor variable: Science knowledge</td>
</tr>
<tr>
<td>( B )</td>
</tr>
<tr>
<td>0.81</td>
</tr>
<tr>
<td>0.14</td>
</tr>
</tbody>
</table>

Using two posttests on science knowledge and general NOS views as the independent variables, the second multiple regression analysis investigated to what extent these variables predicted students’ content-related NOS views as represented by the posttest scores. Preliminary data checking also showed no violation of assumptions in conducting multiple regression analysis. The analysis revealed that the two independent variables accounted for a significant amount of variability in the dependent variable, \( F(2, 80) = 25.37, p < .001 \). The value of 0.388 for R square
indicated that the two predictor variables together accounted for 39% of the variability on content-related NOS posttest scores. In addition, both posttest scores for science knowledge and general NOS views had unique contribution in predicting how well students developed their content-related NOS views as measured in the posttest. The beta weights, presented in Table 4, suggest that science knowledge achievement contributed most to predicting students’ content-related NOS views, followed by general NOS views.

The final multiple regression analysis of the retention test was conducted to assess the simultaneous effects of students’ science knowledge and NOS views on content-related NOS views. Data analysis revealed no serious threats to the assumptions of linearity and normal distribution of residuals of the dependent variable. The result of the multiple regression model found an R of 0.57 and an R square of 0.33. When both science knowledge and general NOS views were treated as predictors, about 33% of the variance in content-related NOS views could be predicted. The overall regression was statistically significant, $F(2, 80) = 19.687, p = 0.00$. Only the science knowledge retention test scores accounted for a significant amount of unique variance in the dependent variable ($p = 0.00$), and general NOS views did not appear to significantly contribute to predicting student content-related NOS views ($p = 0.08$) (see Table 4).

**Discussions and implications**

This research explored how discussing scientific news using different communication approaches influenced student learning, especially the effects that this had on students’ scientific content knowledge, and general and content-related NOS views. Students engaged in either asynchronous or synchronous online discussion analyzed news stories in terms of the accumulative, creative, and community aspects of NOS. Recognizing the potential educational benefits of using scientific news in informal science instruction, the current study combined scientific news with popular social networking tools.

Results showed that all students improved their science content knowledge, and general and content-related NOS views regardless of which group they were assigned. While Johnson (2008) claimed that communication in synchronous and asynchronous approaches had no significant impact on student learning, this study found that students in synchronous groups acquired better understanding of science content knowledge than those in asynchronous groups. The ability to receive quick feedback and responses (Davidson-Shivers, Muilenburg & Tanner, 2001) in synchronous discussion benefited student learning. By providing students with a sense of community (Schwier & Balbar, 2002), synchronous discussion in this study helped students to communicate cognitively and socially.

This study also examined how different types of online discussion approaches influenced student content-related NOS views when they analyzed scientific news in Facebook. When students engaged in online synchronous discussion to analyze the NOS features of scientific news, they understood community aspects of content-related NOS better than those in asynchronous discussion groups. But different communication approaches showed no significant difference in terms of how students analyzed the accumulative and creative features of NOS in scientific news. Like the findings of prior studies (Khishfe, 2008), students in the current study did not show consistent improvement in all aspects of NOS when they received explicit instruction. One reason for this might be the nature of synchronous communication. Schwier and Balbar (2002) emphasized that synchronous strategies such as real time chat experiences created a strong sense of community compared to asynchronous strategies. Therefore, prompting individuals to discuss scientific news synchronously is likely to enhance their scientific community aspect of NOS views as well.

A regression analysis revealed the improvement of students’ general NOS views as well as the significance of science knowledge to foster content-related NOS. As shown in the pretest regression model, students’ general NOS views did not account for their development of content-related NOS views. But the posttest regression showed that general NOS views significantly predicted development of content-related NOS views. Therefore, analyzing scientific news, as presented in the current study, significantly enhanced the development of students’ general NOS views. Moreover, all regression tests revealed that science knowledge significantly predicts student content-related NOS. This finding highlights the value of explicit NOS instruction, strengthening the connection between science knowledge and content-related NOS understanding (Clough, 2006).
We conclude that synchronous online discussion facilitates the development of content knowledge, which fosters understanding of content-related NOS. While scholars believed that asynchronous discussion benefited complex thinking and discussions (Mandernach, Dailey-Hebert, & Donnelly-Sallee, 2007; Romiszowski & Mason, 2004; Schrire, 2006), the findings of this study favored synchronous communication to teach content-related NOS. Such results further indicate that learning NOS, especially when it is embedded within explicit instruction, is not a difficult task for middle school students.

Moreover, consistent with prior studies (Khishfe & Abd-El-Khalick, 2002; Southerland, Gess-Newsome, & Johnston, 2003), the current study highlights how explicit instruction supports NOS learning. The positive learning outcomes reiterate the educational value of engaging individuals in critically reading scientific news in order to learn NOS. While the current research only investigated three aspects of NOS, future studies should continue exploring whether synchronous discussion also benefits other aspects of NOS learning. In addition, it also would be beneficial to explore why synchronous communication can effectively develop students’ science content knowledge and views of community aspects of content-related NOS, but not accumulated knowledge and creativity. Also, this study only analyzed each student’s response to scientific news without considering the context of student conversation from a holistic viewpoint. Future studies should analyze the content of threaded discussion, such as how the process of student interaction with peers influenced student learning.

Acknowledgments

This work was supported mainly by the National Science Council, Taiwan, under Grant Number NSC 98-2511-S009-006-MY3

References


Development of a Survey to Measure Self-efficacy and Attitudes toward Web-based Professional Development among Elementary School Teachers

Chia-Pin Kao1*, Chin-Chung Tsai2, and Meilun Shih3

1Department of Childhood and Education, Southern Taiwan University of Science and Technology, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // 3Center for General Education, National Taiwan University, Taiwan // kcp76@mail.stust.edu.tw // ccttsai@mail.ntust.edu.tw // mshih@ntu.edu.tw

*Corresponding author

(Submitted July 23, 2013; Revised January 2, 2014; Accepted February 4, 2014)

ABSTRACT
The major purpose of this study was to develop a survey to measure elementary school teachers' self-efficacy for web-based professional development. Based on interviews with eight elementary school teachers, three scales of web-based professional development self-efficacy (WPDE) were formed, namely, general self-efficacy (measuring teachers’ general capabilities of using Internet tools or completing basic web-based tasks), interaction self-efficacy (assessing teachers’ perceived abilities of engaging in interactions in online professional development activities), and applying self-efficacy (evaluating teachers’ confidence level of applying what they have learned in web-based professional development to their future teaching). The results of Exploratory Factor Analysis (EFA) indicated that the three scales of WPDE arising from the interviews and the Cronbach alpha values, ranging from 0.92 to 0.85, were acceptable. Subsequently, a large-scale survey was conducted to examine the relationships between teachers’ self-efficacy and attitudes toward web-based professional development with a sample of 214 elementary school teachers in Taiwan using a questionnaire that included WPDE and an established scale for measuring attitudes toward web-based professional development (AWPD). The results indicated that teachers’ self-efficacy played a positive role in their attitudes towards web-based professional development. The results also indicated that teachers’ confidence in their interaction self-efficacy and in their applying self-efficacy was more important than their confidence in their general self-efficacy in predicting their attitudes toward web-based professional development. Implications of the study findings were also discussed.

Keywords
Self-efficacy, Attitude, Web-based professional development

Introduction
Since the pressure of educational reform is constantly increasing, as in-class executors, teachers need to enhance their independence and professional capabilities in order to perform their job with the greatest efficiency (Levy, 2008). The introduction of web-based information technology has begun a trend in education of gaining professional knowledge online. The Internet not only assists teachers greatly in their teaching, but also plays an important role in their learning and professional development. Therefore, they have been encouraged to take part in web-based professional development to meet the demands of their future teaching (Vekiri & Chronaki, 2008). With their feature of no time or space limitations, web-based resources can provide teachers with a more flexible way of pursuing professional development. While enhancing teachers’ willingness to participate in online courses, it can also motivate them to gradually change their teaching from traditional ways (e.g., face-to-face lectures, single location) to a more flexible and adjustable approach based on individual needs (Hartley, 2007).

Self-efficacy is a psychological concept which refers to personal beliefs and expectations regarding whether someone is capable of accomplishing something (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). Internet self-efficacy indicates web users’ self-perceived confidence and expectations of using the Internet. It has been proposed that teachers with high Internet self-efficacy may have a greater chance of success in web-based-related tasks (Kao & Tsai, 2009). Those teachers who possess higher Internet self-efficacy have been found to accept new learning ideas more easily and are usually more willing to try various teaching methods using web-based technology to provide students with different learning experiences (Chen & Tseng, 2012). Therefore, teachers’ self-efficacy in using web-based information technology would influence their willingness to utilize technology in their teaching and/or learning. In other words, teachers’ self-efficacy in applying Internet technology for educational purposes is a key to their participation or engagement in web-based professional development.
Currently, most studies on learners’ self-efficacy focus on traditional learning environments (Clayton, Blumberg, & Auld, 2010). Although there are some studies on learners’ Internet self-efficacy, they generally focus on higher education and often investigate its relationship with college students’ attitudes, beliefs and preferences for online courses (Yang & Tsai, 2008; Yudko, Hirokawa, & Chi, 2008). Thus far, few instruments have been developed to specifically measure self-efficacy for web-based professional development. Since Internet-based programs have become one of the important pathways for elementary school teachers’ professional development, teachers’ web-based professional development self-efficacy is worth further discussion and subsequent study. In order to understand this particular form of self-efficacy, this study explores and evaluates elementary school teachers’ confidence levels regarding their capabilities for web-based professional development.

Moreover, the relationship between self-efficacy and attitudes toward computers and/or the Internet has been examined in some previous studies (Celik & Yesilyurt, 2013; Smarkola, 2008). The results of these studies revealed that learners’ computer attitudes are positively correlated with their computer self-efficacy. Also, learners with greater Internet self-efficacy may have more positive attitudes toward the Internet. While teachers may have more opportunities to learn with web-based professional development, this study aims to explore whether teachers’ self-efficacy might be related to their attitudes in the web-based professional development context.

Therefore, the major purpose of this study was to develop a survey to measure elementary school teachers’ self-efficacy for web-based professional development. Based on semi-structured interviews of elementary school teachers with web-based professional development experience, a survey was first developed to assess teachers’ self-efficacy for web-based professional development. Then, through gathering questionnaire responses from 214 elementary school teachers in Taiwan, the following research objectives were explored in this study:

- Determine the validity of the survey to measure web-based professional development self-efficacy and attitudes toward web-based professional development,
- Determine the web-based professional development self-efficacy expressed by the elementary school teachers,
- Determine the relationships between the teachers’ self-efficacy and their attitudes toward web-based professional development, and
- Use teachers’ web-based professional development self-efficacy to predict their attitudes toward web-based professional development.

**Literature review**

*Internet self-efficacy*

Self-efficacy is a concept introduced by the American psychologist Alberto Bandura. According to Bandura, Barbaranelli, Caprara, and Pastorelli (1996), self-efficacy is a belief that someone has in him/herself of his/her ability to accomplish an action, but which has an influence on the individual’s ways of thinking, motivation, behavior and emotional state of mind. It includes a person’s perceptions of his/her own ability, endeavors, the difficulty level of an assignment, and others’ assistance (Bandura, 2006). While performing learning activities, self-efficacy has influences on learners’ motivation, mission concentration, effort engagement, anxiety or negative thoughts of self-denial (Kinzie, 1990; Semiatin & O’Connor, 2012). Studies on the relationships between self-efficacy and learning effectiveness have found that students with higher self-efficacy are more capable of cognition and metacognition, while being more persevering in the learning context (Burgoon, Meece, & Granger, 2012; McCombs, 1984). In other words, self-efficacy can show the confidence perceived by a learner and therefore can greatly influence his/her motivation and learning results.

Studies have found that learners’ self-efficacy could predict their academic performance in conventional learning environments (e.g., Zhu, Chen, Chen, & Chern, 2011). Subsequently, researchers have applied the self-efficacy concept to Internet-related situations, and proposed Internet self-efficacy as a self-evaluation for Internet usage and execution, such as receiving e-mails, downloading/uploading files via ftp servers, using search engines, and taking part in online discussions (Hsu & Chiu, 2004; Kim & Glassman, 2013).

Tsai, Chuang, Liang and Tsai’s (2011) work reviewed some studies relating to self-efficacy in Internet-based learning environments. Their results showed that computer self-efficacy, Internet self-efficacy and Internet-Based Learning self-efficacy were consistent with the original concept of the self-efficacy theory. On the basis of their
findings, the definitions for self-efficacy regarding computers, the Internet and web-based professional development are as follows:

- **Computer Self-Efficacy (CSE)** can be considered a domain specific measure of self-efficacy that reflects a person’s belief in his/her ability to perform specific computer tasks.
- **Internet Self-Efficacy (ISE)** examines learners’ confidence in their general skills or knowledge of operating Internet functions or applications in the Internet-based learning condition.
- **Web-based professional development self-efficacy (WPDSE)** explores teachers’ confidence in their participation in and their expected performance of using web-based professional development.

Obvious connections have been found between past Internet experience, outcome expectations and Internet self-efficacy (Wu & Tsai, 2006). Studies have proven that if students’ self-efficacy is enhanced, not only will the frequency of their Internet use be increased, but their performance will also be improved (Burgoon et al., 2012). These related studies have not only extended the results of Bandura’s study, but have also enhanced the applicability and value of the concept of self-efficacy to the use of technology.

**Web-based professional development self-efficacy**

In the age of Internet technology, the way in which teachers undertake professional development has changed from traditional face-to-face, single-location training to developmental activities in advanced web-based environments. Because professional development is undertaken via the Internet, teachers can choose courses according to the availability of place and time (Chen, Chen, & Tsai, 2009; Kinzie, Whitaker, Neesen, Kelley, Matera, & Pianta, 2006). Meanwhile, teachers with the same interests and expertise can make connections and form a network to share their experiences of professional development and teaching resources. Therefore, web-based professional development courses not only enrich teachers’ professional knowledge and experience, but can also help them to cross school boundaries.

Some studies have pointed out that teachers with higher technology-related self-efficacy have increased inclinations to use technology in their teaching (Paraskeva, Bouta, & Papagianni, 2008). These results suggest that when teachers have higher confidence in using information technology, their willingness to apply and integrate it in their teaching is increased as well.

Although some studies discuss self-efficacy in educational environments, most discussions of web-based learning are related to students in senior high schools or universities, and indicate that in web-based learning environments, learners’ web-related self-efficacy not only has an influence on their information usage, but is also a driving factor in the continuous pursuit of learning (Tsai, 2009). Similar to the important role of students’ Internet self-efficacy in their web-based learning, teachers’ web-based professional development self-efficacy may also play a significant role in their intention to engage in and their participation in web-based professional development. By the same token, teachers’ web-based professional development self-efficacy is an important issue for investigation. However, there is still little research in this area. Therefore, it is appropriate to develop and validate an instrument that seeks to examine teachers’ self-efficacy of web-based professional development, that is, their confidence levels in their capabilities of undertaking such professional development.

**The relationship between teachers’ self-efficacy and attitudes toward web-based professional development**

The relationship between attitudes toward and self-efficacy regarding computers, the Internet and Internet-based learning has been examined in a number of studies (e.g., Liang, Wu, & Tsai, 2011; Susskind, 2008). It has been found that learners’ computer attitudes are positively correlated with their computer self-efficacy (Pamuk & Peker, 2009). Also, learners with greater Internet self-efficacy may have more positive attitudes toward web-based professional development (Kao & Tsai, 2009). Expanding on the Theory of Planned Behavior (TPB), Ajzen (2002) proposed that behavioral intentions are influenced by attitudes and self-efficacy. It has been proven that beliefs and attitudes are the key determinants of both initial technology usage and long-term usage intentions and behaviors (Sun, 2008). In light of the above research, it can be concluded that a learner’s self-efficacy and attitudes toward technology are consistently and strongly related to their usage of the technology. Therefore, the investigation of
teachers’ self-efficacy and attitudes toward web-based professional development can provide insights into their web-based professional development intentions as well as their performance.

It has been shown that successful web-based professional development depends on teachers’ Internet self-efficacy and attitudes toward its learning environment (Kao & Tsai, 2009). In some studies, teachers’ self-efficacy and attitudes have been identified as important variables that are related to their use of technology (Pynoo et al., 2011). Consequently, our study suggests that there are some relationships between teachers’ self-efficacy and attitudes toward web-based professional development. If more knowledge about these relationships can be acquired, researchers can know more about how to effectively improve web-based professional development performance and how to help teachers adopt web-based professional development.

Method

In order to explore and evaluate elementary school teachers’ web-based professional development self-efficacy, the research design of this study consisted of two procedures which combined both qualitative and quantitative methods of data collection and analysis. First, semi-structured interviews were carried out at the beginning of the study with the purpose of gathering information for the construction of the questionnaire. Then, the researchers developed a questionnaire on the basis of the interview data, and a large-scale survey was subsequently administered.

Sample

Convenience sampling was used to gather teachers’ responses about their web-based professional development self-efficacy. Eight elementary school teachers were individually interviewed in this study. The interview data were used to explore what general perceptions, expectations, and experiences the teachers held regarding web-based professional development. In addition, the researchers also used content derived from the interview data to construct a questionnaire for exploring teachers’ web-based professional self-efficacy. The interviewed teachers came from eight elementary schools in Taiwan. Six of them were females, and two were males. All of them had prior experience of participating in web-based professional development in the past two years.

In the second part of the study, a total of 214 teachers from 20 elementary schools in Taiwan were selected as the subjects. All of the participating teachers had actual experience of web-based professional development. Though the sample came from a convenient sample, these teachers were from different geographical areas in Taiwan and across schools of varying socio-economic status.

Data collection and analysis

Two types of data were collected in this study. First, eight elementary school teachers with web-based professional development experience were invited to participate in semi-structured interviews. The interview questions aimed to provide a better understanding of the teachers’ perceptions, expectations, and experiences of web-based professional development. Example questions included:

- What do you think about web-based professional development? How confident do you feel when you participate in web-based professional development courses?
- What are the advantages and disadvantages of participating in web-based professional development programs? Are these related to your confidence in taking part in web-based professional development? In what ways?
- What kinds of problems did you encounter in web-based professional development? How did you solve the problems? How confident are you about solving these problems?
- Would you recommend that other teachers attend web-based professional development programs?

Each participating teacher was interviewed individually by one researcher for about forty minutes. All interviews were audio-recorded and fully transcribed to provide detailed information regarding the teachers’ ideas about their web-based professional development self-efficacy. The verbatim transcripts were read and discussed by the researchers. Important sentences or keywords that represented the teachers’ web-based professional development
self-efficacy were marked. Then, similarities and differences in the marked sentences or keywords across each teacher were explored and summarized as the basis for creating the questionnaire items.

Secondly, the final sample in this study included 214 elementary school teachers. They were used convenience sampling from 20 elementary schools from both northern and southern Taiwan. Then, an exploratory factor analysis (EFA) with principal components analysis and Varimax rotation with Kaiser normalization were employed to validate the structure of the questionnaire using the SPSS 12 software. Based on factors resulting from the EFA analyses, and the responses to the questionnaire items, the teachers’ levels of web-based professional development self-efficacy were determined using means and standard deviations. Correlation analysis was also carried out to examine the relationship between teachers’ self-efficacy and their attitudes toward web-based professional development. Then, through a stepwise multiple regression analysis, teachers’ web-based professional development self-efficacy was tested as a predictor to explain their web-based professional development attitudes.

**Questionnaire**

The questionnaire in this study comprised of two main parts. The first part consisted of the Web-based Professional Development Self-Efficacy (WPDSE) questionnaire developed based on the interviews with the eight elementary school teachers, using prior studies (e.g., Kao & Tsai, 2009) as a reference framework to measure the teachers’ self-evaluation of their confidence level and capabilities of engaging in web-based professional development. The researchers went through the interview transcripts to identify the main themes and keywords for the construction of the initial items of the WPDSE, which included items measuring elementary school teachers’ general abilities (e.g., uploading or downloading files from the web) to engage in web-based professional development. Furthermore, items about the most current web-based technologies such as blogs and social networks were also added as WPDSE items to reflect teachers’ possible technology usage nowadays. Moreover, the questionnaire also included items for evaluating the confidence level of applying what the teachers had learned in their web-based professional development to their teaching practice, including teaching performance and integrating technologies in their classes. As a result, WPDSE included three scales (general self-efficacy, interaction self-efficacy, and applying self-efficacy), each consisting of seven to eight items, giving a total of 22 items in the initial version. All items were presented with a five-point Likert scale, ranging from one “strongly disagree” to five “strongly agree.” The details of the three scales are as follows:

- **General self-efficacy**: measuring teachers’ general capabilities of using Internet tools or completing basic web-based tasks. This scale focuses on teachers’ confidence in their basic ability to use web tools.

- **Interaction self-efficacy**: assessing teachers’ perceived abilities of engaging in interactions in online professional development activities. This scale focuses on teachers’ confidence in their ability to interact with others or to fulfill web-based tasks through interactions.

- **Applying self-efficacy**: evaluating teachers’ confidence level of applying what they have learned in web-based professional development to their teaching. This scale focuses on teachers’ confidence in applying what they have learned in web-based professional development to their teaching practice.

The second part of the questionnaire was the Attitudes toward Web-Based Professional Development (AWPD), based on Kao and Tsai’s (2009) research. The AWPD implemented in this study included five scales and a total of 27 items. The details of the five scales are as follows:

- **Perceived usefulness scale**: assessing perceptions of the extent to which teachers perceive that the impact of web-based professional development is positive and useful. That is, the higher the score, the stronger the teachers’ beliefs in the usefulness of web-based professional development.

- **Perceived ease of use scale**: assessing the extent to which teachers perceive that web-based professional development is easy to use. That is, higher scores indicate higher agreement with the ease of use of web-based professional development.

- **Affection scale**: measuring perceptions of the extent to which teachers express favorable feelings about web-based professional development. That is, higher scores suggest more positive feelings about web-based professional development.
• Anxiety scale: measuring perceptions of the extent to which teachers experience anxiety about web-based Professional development. This scale is scored in reverse. Thus, higher scores indicate less anxiety about web-based professional development.

• Behavior scale: measuring perceptions of the extent to which teachers perceive actual practice and express willingness to use web-based professional development. That is, the higher the score, the more willing the teachers are to use web-based professional development.

Kao and Tsai’s (2009) study reported alpha reliability coefficients for the five scales as 0.92 (perceived usefulness), 0.92 (perceived ease of use), 0.87 (affection), 0.88 (anxiety) and 0.93 (behavior). The alpha value for the overall AWPD was 0.91, and these factors explained 80.65% of the total variance. Therefore, these scales were deemed to be sufficiently reliable for assessing teachers’ attitudes toward web-based professional development.

Results

Validation of WPDSE and AWPD

This study utilized exploratory factor analysis to clarify the structure of the perceptions of web-based professional development based on the data obtained from the large scale survey. Items with a factor loading of less than 0.5 were subject to deletion from the item pool. In addition, Cronbach’s alpha values were estimated to confirm the reliability of the overall instrument and each item. Consequently, teachers’ web-based professional development self-efficacy was grouped into three scales: general self-efficacy, interaction self-efficacy and applying self-efficacy. Thus, using exploratory factor analysis, the initial 22 items were all retained and were in the same scales as indicated by the interviews (as shown in Table 1).

Table 1. Rotated factor loadings and Cronbach alpha values for the web-based professional development self-efficacy scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1: General self-efficacy, ( \alpha = 0.94 ), Mean = 4.57, ( SD = 0.49 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE 1</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE 2</td>
<td>0.86</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>GSE 3</td>
<td>0.84</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>GSE 4</td>
<td>0.82</td>
<td>0.79</td>
<td>0.81</td>
</tr>
<tr>
<td>GSE 5</td>
<td>0.79</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>GSE 6</td>
<td>0.78</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td>GSE 7</td>
<td>0.77</td>
<td>0.74</td>
<td>0.81</td>
</tr>
<tr>
<td>GSE 8</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Factor 2: Interaction self-efficacy, \( \alpha = 0.94 \), Mean = 3.94, \( SD = 0.65 \) |          |          |          |
| ISE 1                      |          | 0.81     |          |
| ISE 2                      |          | 0.81     |          |
| ISE 3                      |          | 0.79     |          |
| ISE 4                      |          | 0.78     |          |
| ISE 5                      |          | 0.77     |          |
| ISE 6                      |          | 0.75     |          |
| ISE 7                      |          | 0.74     |          |

| Factor 3: Applying self-efficacy, \( \alpha = 0.92 \), Mean = 4.05, \( SD = 0.48 \) |          |          |          |
| ASE 1                      |          |          | 0.82     |
| ASE 2                      |          |          | 0.82     |
| ASE 3                      |          |          | 0.81     |
| ASE 4                      |          |          | 0.81     |
| ASE 5                      |          |          | 0.78     |
| ASE 6                      |          |          | 0.74     |
| ASE 7                      |          |          | 0.74     |
| Percentage of variance     | 45.98    | 15.22    | 9.16     |
| Overall \( \alpha = 0.95 \) |          |          |          |
| Total variance explained   | 70.35%   |          |          |
The alpha values for the three scales were 0.94 (general self-efficacy, 8 items), 0.94 (interaction self-efficacy, 7 items) and 0.92 (applying self-efficacy, 7 items). The overall alpha value was 0.95, and the factors explained 70.35% of total variance. By and large, the study revealed acceptably high (Thompson & Daniel, 1996) alpha reliability coefficients for all items. Thus, the scales were considered to be sufficiently reliable for assessing the teachers’ web-based professional development self-efficacy. The 22 items for final inclusion in WPDSE are listed in Appendix 1.

Using exploratory factor analysis, the initial 27 items of AWPD were reduced to 18 items in five scales (as shown in Table 2). The final scales were the same as those proposed in Kao and Tsai’s (2009) study, namely, perceived usefulness, perceived ease of use, affection, anxiety and behavior. The reliability coefficients for the five scales of the AWPD were 0.87 (perceived usefulness, 4 items), 0.87 (perceived ease of use, 4 items), 0.86 (affection, 3 items), 0.87 (anxiety, 3 items) and 0.91 (behavior, 4 items). The alpha value of the whole AWPD questionnaire is 0.87, and these factors explained 77.35% of total variance. Therefore, these scales were deemed to have sufficient internal consistency for assessing teachers’ attitudes toward web-based professional development.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1: Perceived ease of use, $\alpha = 0.87$, Mean = 4.17, $SD = 1.04$</td>
<td>Ease 1: 0.87</td>
<td>Ease 2: 0.85</td>
<td>Ease 3: 0.86</td>
<td>Ease 4: 0.87</td>
<td></td>
</tr>
<tr>
<td>Factor 2: Behavior, $\alpha = 0.91$, Mean = 3.97, $SD = 1.28$</td>
<td>Behavior 1: 0.90</td>
<td>Behavior 2: 0.87</td>
<td>Behavior 3: 0.87</td>
<td>Behavior 4: 0.83</td>
<td></td>
</tr>
<tr>
<td>Factor 3: Perceived usefulness, $\alpha = 0.87$, Mean = 4.34, $SD = 0.87$</td>
<td>Perceived use 1: 0.77</td>
<td>Perceived use 2: 0.81</td>
<td>Perceived use 3: 0.83</td>
<td>Perceived use 4: 0.76</td>
<td></td>
</tr>
<tr>
<td>Factor 4: Anxiety, $\alpha = 0.87$, Mean = 3.57, $SD = 1.48$</td>
<td>Anxiety 1: 0.87</td>
<td>Anxiety 2: 0.77</td>
<td>Anxiety 3: 0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 5: Affection, $\alpha = 0.86$, Mean = 4.06, $SD = 0.96$</td>
<td>Affection 1: 0.86</td>
<td>Affection 2: 0.82</td>
<td>Affection 3: 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>44.64</td>
<td>13.47</td>
<td>7.49</td>
<td>6.12</td>
<td>5.63</td>
</tr>
</tbody>
</table>

Overall $\alpha = 0.87$. Total variance explained is 77.35%.

### Elementary school teachers’ web-based professional development self-efficacy

The results indicated that the teachers scored highly on each of the three WPDSE scales, with an average of 4.57 per item on the general self-efficacy scale, an average of 3.94 per item on the interaction self-efficacy scale, and an average of 4.05 per item on the applying self-efficacy scale on the 1-5 Likert measurement. This result indicates that the teachers tended to display greater confidence in using the Internet for general purposes, and showed high confidence in applying learning outcomes to their teaching practice after taking part in web-based professional development. However, their relatively low score on the interaction self-efficacy scale still implies that some teachers expressed comparatively less confidence in engaging interactions during web-based professional development.
Relationships between self-efficacy and attitudes toward web-based professional development

This study also examined the relationship between teachers’ web-based professional development self-efficacy (WPDSE) and their attitudes toward web-based professional development (AWPD), as shown in Table 3.

The relationships between the WPDSE and the AWPD indicated that all of the variables were significantly positively correlated with each other (all $r > 0.19, p < 0.01$), except that no statistical correlation was found between applying self-efficacy and anxiety. These results, in general, support the argument that teachers expressing higher web-based professional development self-efficacy display more positive perceptions of, better affection for, less anxiety about, and higher willingness to participate in web-based professional development. In particular, the teachers’ responses on the usefulness scale were relatively more highly correlated with those on the applying self-efficacy scale ($r = 0.65, p < 0.01$). It seems that teachers with higher self-efficacy in applying learning outcomes to their teaching practice with web-based technology might tend to perceive web-based professional development as being useful.

By and large, the teachers’ scores on the interaction self-efficacy scale and applying self-efficacy were fairly highly correlated with all of the scales of the AWPD, with the exception of the correlation between applying self-efficacy and anxiety. This indicates that teachers with higher confidence in engaging in interaction in online learning activities and using WPD for teaching purposes tend to express more positive perspectives, better affection, less anxiety and greater willingness to use WPD. This also suggests that, for enhancing teachers’ attitudes toward web-based professional development, teachers’ interaction self-efficacy and applying self-efficacy are more important than their general self-efficacy for Internet-related functions.

Table 3. Correlation of the teachers’ self-efficacy and attitudes toward web-based professional development

<table>
<thead>
<tr>
<th>Scale</th>
<th>Usefulness</th>
<th>Ease of use</th>
<th>Affection</th>
<th>Anxiety*</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>General self-efficacy</td>
<td>0.19**</td>
<td>0.37***</td>
<td>0.23***</td>
<td>0.28***</td>
<td>0.27***</td>
</tr>
<tr>
<td>Interaction self-efficacy</td>
<td>0.43***</td>
<td>0.57***</td>
<td>0.56***</td>
<td>0.31***</td>
<td>0.54***</td>
</tr>
<tr>
<td>Applying self-efficacy</td>
<td>0.65***</td>
<td>0.55***</td>
<td>0.59***</td>
<td>0.10</td>
<td>0.51***</td>
</tr>
</tbody>
</table>

**$p < 0.01$.

* a higher score indicates less anxiety about web-based professional development

Predicting teachers’ attitudes toward web-based professional development by self-efficacy

While teachers may have more opportunities to learn with web-based professional development, it is plausible to hypothesize that teachers’ web-based professional development self-efficacy may be related to their attitudes toward web-based professional development. Hence, this study also examined the predictive power of teachers’ web-based professional development self-efficacy to predict their attitudes toward web-based professional development, as shown in Table 4.

Table 4. Multiple regression models of predicting teachers’ attitudes toward web-based professional development (n=214)

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Predicting variables</th>
<th>B</th>
<th>S.E.</th>
<th>$\beta$</th>
<th>$T$</th>
<th>$R^2$ (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>ASE</td>
<td>0.55</td>
<td>0.06</td>
<td>0.58</td>
<td>9.50***</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>ISE</td>
<td>0.09</td>
<td>0.04</td>
<td>0.13</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>1.39</td>
<td>0.21</td>
<td>0.39</td>
<td>6.71***</td>
<td></td>
</tr>
<tr>
<td>Ease of use</td>
<td>ISE</td>
<td>0.37</td>
<td>0.06</td>
<td>0.39</td>
<td>6.25***</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>ASE</td>
<td>0.46</td>
<td>0.08</td>
<td>0.35</td>
<td>5.71***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.66</td>
<td>0.29</td>
<td>0.22</td>
<td>2.28*</td>
<td></td>
</tr>
<tr>
<td>Affection</td>
<td>ASE</td>
<td>0.56</td>
<td>0.08</td>
<td>0.43</td>
<td>7.11***</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>ISE</td>
<td>0.41</td>
<td>0.07</td>
<td>0.42</td>
<td>6.14***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GSE</td>
<td>0.20</td>
<td>0.08</td>
<td>0.16</td>
<td>2.49*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>0.87</td>
<td>0.36</td>
<td>0.16</td>
<td>2.44*</td>
<td></td>
</tr>
<tr>
<td>Anxiety*</td>
<td>ISE</td>
<td>0.24</td>
<td>0.11</td>
<td>0.22</td>
<td>2.82</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>GSE</td>
<td>0.23</td>
<td>0.10</td>
<td>0.16</td>
<td>1.98*</td>
<td></td>
</tr>
</tbody>
</table>
Table 4 shows that teachers with stronger applying self-efficacy ($t = 7.11, p < .001$), interaction self-efficacy ($t = 6.14, p < .001$) and general self-efficacy ($t = 2.49, p < .05$) expressed more favorable affection toward web-based professional development, explaining the relationship for the $R$ of 45%. Teachers with higher interaction self-efficacy ($t = 2.82, p < .01$) and general self-efficacy ($t = 1.98, p < .05$) displayed less anxiety toward web-based professional development, explaining the relationship for the $R$ of 11%. Moreover, the result of this study found that interaction self-efficacy and applying self-efficacy could significantly and positively predict perceived usefulness, perceived ease of use, and behavior, respectively explaining 43%, 41%, and 36% of teachers’ attitudes toward web-based professional development. This indicates that those teachers with higher confidence in using WPD for learning and teaching purposes tended to express more positive perceptions of, better feelings about, and greater willingness to use WPD. In particular, teachers with high interaction self-efficacy displayed a positive effect on perceived ease of use, anxiety and behavior, indicating that interaction self-efficacy may play the most important role in the AWPD, as it was positively associated with all aspects of their attitudes toward web-based professional development. Applying self-efficacy was also a significantly positive predictor of the perceived usefulness, ease of use, affection, and behavior of the AWPD.

In sum, the results shown in Table 4 revealed the importance of interaction self-efficacy and applying self-efficacy in predicting AWPD, indicating that teachers with positive interaction self-efficacy and applying self-efficacy of web-based professional development are more likely to have favorable attitudes toward web-based professional development. However, general self-efficacy can only predict affection and anxiety for AWPD.

**Discussion and conclusions**

Web-based professional development self-efficacy is a pivotal construct for understanding a wide range of online learning activities for teachers. By assessing teachers’ web-based professional development self-efficacy, researchers can acquire some indications of their expected outcomes derived from web-based.

The WPDSE and the AWPD administered in this study are deemed to be sufficiently reliable for assessing teachers’ self-efficacy and their attitudes toward web-based professional development, respectively. By means of these two instruments, the present study explored a group of Taiwanese elementary school teachers’ self-efficacy and their attitudes toward web-based professional development. The results of this study derived from the correlation analyses demonstrate that the teachers’ web-based professional development self-efficacy was positively correlated with their attitudes toward web-based professional development. Teachers with higher web-based professional development self-efficacy expressed favorable attitudes toward web-based professional development. Also, the correlational analyses indicated that teachers who had stronger interaction self-efficacy and applying self-efficacy showed more positive perceptions of, stronger liking for, and stronger willingness to participate in web-based professional development.

Furthermore, the regression analysis revealed that teachers’ interaction self-efficacy was the most significant positive predictor (i.e., the first-selected factor for the step-wise regression model) for the ease of use, anxiety and behavior of web-based professional development. Teachers’ positive expectations of participating and engaging in interactions in online learning activities are critical for their perceptions of accessibility, lower anxiety and behavior to take part in web-based professional development. In addition, the regression analysis also revealed that teachers’ applying self-efficacy was also the main significant positive predictor (i.e., the first-selected factor for the step-wise regression model) of the usefulness and affection of web-based professional development. This seems to show that teachers’ confidence in using WPD for teaching purposes is critical for their positive perceptions of usefulness and favorable feelings toward web-based professional development activities. The findings above are also similar to those in the previous studies concerning the influences of self-efficacy on attitudes regarding the computer and/or the Internet and/or the web-based learning (Kao & Tsai, 2009; Susskind, 2008; Wu & Tsai, 2006). That is, enhancing stronger self-efficacy in using WPD of engaging interaction and applying for teaching purposes may help teachers shape...
better attitudes toward web-based professional development. Thus, the importance of WPDSE should be highlighted for teachers’ web-based professional development.

The finding above is worth noting as it implies that teachers’ Internet self-efficacy may not be the most significant factor to predict their attitudes toward web-based professional development. While predicting teachers’ attitudes toward web-based professional development, their confidence in actually using WPD (i.e., their interaction self-efficacy and applying self-efficacy) is more important than their confidence in using Internet-related functions (i.e., their general self-efficacy). Therefore, educators should try to find effective ways to improve teachers’ perceptions of engaging in interaction in online learning activities and their capacity to use technology-related teaching methods in the classroom. Then, teachers’ attitudes toward web-based professional development can be enhanced accordingly.

The present results indicate that the government, educational administrators and instructional designers would benefit from being more attentive to teachers’ percepts of web-based professional development self-efficacy. Likewise, our results provide encouraging support for a conceptualization of web-based professional development self-efficacy at the teacher level and for a new measure of this construct. This study may add to our understanding of the nature of web-based professional development behaviors by teachers in educational practice. As Shea and Bidjerano (2010) pointed out, learners’ learning performance can be altered successfully by interventions which increase their beliefs in their competence and their confidence. Besides, Markauskaite (2007) has suggested that an Internet-related enhancement program that was implemented might have resulted in successful mastery experiences for the teachers. Thus, given that useful and specific training programs seem to be practicable for elementary school teachers, they may be helpful for enhancing teachers’ professional development and teaching practice.

One limitation of this study is that, technically, the use of questionnaires for examining self-efficacy and attitudes toward web-based professional development may suffer from a lack of strong validity. To further explore teachers’ self-efficacy of and attitudes toward web-based professional development, researchers can place teachers in a real web-based learning environment. Through gathering extensive data when they are experiencing the web-based learning activities, researchers may know more about teachers’ real perceptions of this new type of learning.

Finally, in light of some previous research (Susskind, 2008; Pynoo et al., 2011; Pamuk & Peker, 2009) which underscored the effect of Internet or computer self-efficacy on increasing attitude levels, our results showed that a similar effect was observed with our sample. Indeed, the result of this study is helpful in facilitating an understanding of teachers’ web-based professional development self-efficacy. By using the WPDSE questionnaire, researchers are encouraged to conduct follow-up studies for different school levels as well as various areas (such as urban or rural) to acquire a better understanding of the possible differences in teachers’ characteristics in terms of web-based professional development self-efficacy. In sum, this study provides some innovative thoughts about elementary school teachers’ web-based professional development self-efficacy, and the practical importance of the results is noted, as they can help the government and educational administrators understand teachers’ perceptions and behaviors, and adjust their professional development policies accordingly. With the questionnaires, follow-up investigations can be put into practice to examine the interplay among beliefs, preferences and learning or teaching behaviors in the web-based context. Such studies will help to evaluate and predict the efficiency of web-based professional development, and furthermore, can make comparisons between web-based and conventional situations.

**Acknowledgments**

Funding of this research work is supported by National Science Council, Taiwan, by grant numbers 101-2511-S-218-006.

**References**


Appendix

Appendix 1: The questionnaire items on the Web-based Professional Development Self-Efficacy questionnaire (WPDSE) survey (final version)

General self-efficacy
I feel confident about using a Web browser like “Internet Explorer” or “Firefox.”
I feel confident about reading the content from the Web.
I feel confident about clicking the hyperlink to connect to another Website.
I feel confident about keying in the Website address to connect to a particular Website.
I feel confident about printing out the content of a Website.
I feel confident about copying images or text on the Web into the WORD software.
I feel confident about searching for information on the Web using keywords.
I feel confident about uploading or downloading files from the Web.

Interaction self-efficacy
I feel confident about selecting appropriate web-based professional development courses.
I feel confident about registering for web-based professional development courses.
I feel confident about reading the contents in web-based professional development courses.
I feel confident about interacting with teachers in web-based professional development courses.
I feel confident about asking or answering questions in web-based professional development courses.
I feel confident about completing assigned course work in web-based professional development courses.
I feel confident about searching for relevant information for web-based professional development courses on the Web.

Applying self-efficacy
After attending web-based professional development, I feel confident about enhancing my teaching performance.
After attending web-based professional development, I feel confident about enriching my course contents.
After attending web-based professional development, I feel confident about applying multiple teaching strategies in my classes.
After attending web-based professional development, I feel confident about extending my teaching resources.
After attending web-based professional development, I feel confident about integrating technologies in my teaching.
After attending web-based professional development, I feel confident about enhancing students’ learning motivations.
After attending web-based professional development, I feel confident about looking for appropriate web resources to guide my students’ learning.

Appendix 2: The questionnaire items on the attitudes toward web-based professional development (AWPD) survey (final version)

Perceived usefulness
1. Web-based professional development helps my instruction become more interesting.
2. Web-based professional development helps to increase my creativity for instruction.
3. Web-based professional development effectively enhances my learning.
4. Web-based professional development improves my professional knowledge.

Perceived ease of use
1. It is easy for me to use web-based professional development on the Internet.
2. It is convenient to receive training on the job by using web-based professional development.
3. The content of web-based professional development is clear, and easy to access for learning.
4. The learning of web-based professional development is flexible.

Affection
1. I think it is interesting to use web-based professional development.
2. Web-based professional development provides an interesting and attractive environment.
3. Using web-based professional development can improve my teaching ability.
Anxiety
1. Using web-based professional development makes me feel anxious.
2. Using web-based professional development makes me feel uncomfortable.
3. Using web-based professional development is boring.

Behavior
1. I hope to spend more time using web-based professional development.
2. I want to increase my use of web-based professional development in the future.
3. I would be glad to use web-based professional development in the future.
4. I will recommend the use of web-based professional development to others.
A Problem-based Ubiquitous Learning Approach to Improving the Questioning Abilities of Elementary School Students

Pi-Hsia Hung1, Gwo-Jen Hwang2*, Yueh-Hsun Lee1, Tsung-Hsun Wu1, Bahtijar Vogel3, Marcelo Milrad3 and Emil Johansson3

1Department of Education, National University of Tainan, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // 3Department of Media Technology, Linnaeus University, Sweden // hungps@mail.nutn.edu.tw // gihwang.academic@gmail.com // asdgary.asd@yahoo.com.tw // tnwuts@tn.edu.tw // bahtijar.vogel@lnu.se // marcelo.milrad@lnu.se // emil.johansson@lnu.se

*Corresponding author

(Submitted July 25, 2013; Revised November 10, 2013; Accepted February 1, 2014)

ABSTRACT

The purpose of this study is to investigate the effects of a ubiquitous problem-based learning system (UPBLS) on students’ question-raising performance in field inquiry activities. An experiment was conducted on an elementary school natural science course. A total of 43 sixth and fifth grade elementary students divided into experienced and novice groups participated in three field observation activities with on-line discussion over a period of seven months to evaluate the changes in their questioning ability. Moreover, a rubric for evaluating the students’ questioning ability was developed and validated. Supported by UPBLS, the students collected the required wetland ecology data in three wetland field observation activities. The experimental results show that the students’ questioning abilities significantly improved during the learning process. Moreover, it was found that both the experienced and the novice students had similar progress trends, implying that the proposed approach is helpful for improving the questioning abilities of students with different levels of prior knowledge.

Keywords

Scientific inquiry, Ubiquitous learning, Problem-based learning, Questioning ability

Introduction

Questioning has been recognized as being an important ability for scientific inquiry and knowledge construction (Scardamalia, 2002; Tan & Seah, 2011). Educators have indicated that fostering students’ scientific questioning skills needs to be heightened in the 21st century since the impact of science and technology significantly affects many aspects of our daily lives (Tan & Seah, 2011). That is, sufficient and proper practice of asking questions could possibly lead students into a positive cycle that enhances their questioning ability, which is helpful to them in terms of improving their domain knowledge (Scardamalia, 2002). Therefore, it is important to situate students in a questioning practicing environment that provides rich information for them to make observations and investigations in order to solve problems.

Problem-based learning (PBL), which focuses on spontaneity, collaboration, and flexible problem-solving skills, is such an approach that engages students in problem-solving scenarios. In the past decades, PBL has become increasingly popular in settings from K-12 to undergraduate education (Barrows, 2000; Dochy et al., 2003; Gallagher et al., 1992; Hmelo, Holton, & Kolodner, 2000; Torp & Sage, 2002; Williams & Hmelo, 1998). According to the definition of Hung, Jonassen and Liu (2008), PBL is an instructional approach that initiates students’ learning by creating a need to solve an authentic problem. During the problem solving process, students construct domain knowledge and develop both problem-solving skills and self-directed learning skills while working toward a solution to a problem. A number of researchers have confirmed the benefits and effectiveness of PBL (Dolmans, Schmidt, & Gijselaers, 1995; Evenson & Hmelo, 2000; Hmelo-Silver, 2004).

In the past decade, the popularity of computer and information technology has further enabled PBL to be applied to various learning areas with different learning supports (Lu, Lajoie, & Wiseman, 2010; Resta & Laferrière, 2007; Rienties et al., 2012). In recent years, owing to the speedy advancement and popularity of wireless communication and mobile technologies, ubiquitous learning environments that integrate real-world and digital-world resources and learning scenarios have provided a new opportunity for implementing technology-enhanced PBL activities (Sharple,
Milrad, Arnedillo-Sánchez, & Vavoula, 2009). This innovative pedagogical method, which is defined as ubiquitous problem-based learning (UPBL), has been confirmed as a potential and productive learning approach (Hung et al., 2012). Without the constraints of a physical space or specific time for learning, ubiquitous learning further strengthens the superiority of PBL. Among the learning areas to which the technology-enhanced PBL activities have been applied, science is one of the most general and suitable subjects, especially for conducting inquiry-based activities in the field (Hung, Lin, & Hwang, 2010).

However, one of the key elements of solving problems, that is, students’ questioning ability, has seldom been investigated. Questioning ability, referring to the skills of exploring environments, understanding contexts, organizing information, and finally proposing a valuable and answerable driving question, has been recognized as an important component of scientific literacy and the cornerstone of scientific inquiry (Chin, 2002; Hofstein, Novon, Kipnis, & Mamlok-Naaman, 2005). Researchers have pointed out that computer and network technologies are likely to provide a good alternative for improving the quality of questioning since students might be more willing to express their opinions and raise questions in technology-based environments than in traditional classrooms (Hu & Chiou, 2012). Therefore, it could be worthwhile to apply the UPBL approach to improving the questioning ability of students.

In addition to technology-based environments, students’ experience is also the main variable affecting the learning progress within inquiry activities. An experienced student, compared to a novice, might be defined as someone who has spent many hours training or solving problems in inquiry learning, and has acquired more knowledge that affects what they notice, the information they remember and recall, as well as how they reason and solve problems (Bransford, Brown, & Cocking, 2000; Petcovic & Libarkin, 2007). Without considering students’ inquiry experience, a learning approach may only benefit some of the students. However, few studies have taken students’ experience into account when proposing systems or approaches for supporting scientific inquiry activities. Therefore, it is worth investigating whether there is a significant difference between the questioning ability progress of the experienced and novice learners in the inquiry activities using the UPBL.

In the meantime, researchers have emphasized the importance of situating students in authentic learning environments (Brown, Collins, & Duguid, 1989; Hwang, Wu, Zhuang, & Huang, 2013), and have indicated the potential of using mobile, wireless communication and sensing technologies (such as QR-codes or Global Positioning Systems) in providing learning supports to students in real-world explorations (Hwang, Tsai, & Yang, 2008). Along these lines, the purpose of this present study is to build up a ubiquitous learning platform for students, and investigate the effect of promoting questioning ability in the problem-based scientific inquiry activities. Moreover, a scoring rubric was developed, which played the important role of guiding and encouraging the students to propose quality questions in the field trip as well as evaluating their questioning ability (Creswell, 2009; Neuman, 2004; Fan & Lê, 2011). Accordingly, the following research questions are investigated:

- Do the rubrics used to assess questioning ability have reasonable reliability and validity?
- Can the UPBL approach improve students’ questioning ability in the inquiry activities?
- Is there a significant difference between the questioning ability progress of the experienced and novice learners in the inquiry activities using the UPBL?

Literature review

Questioning ability

Researchers have indicated the importance of the role of questioning ability in students’ scientific inquiry and knowledge building performance (Chin & Kayalvizhi, 2002; Scardamalia, 2002; Tan & Seah, 2011). Similarly, the ability to ask good questions is also regarded as an essential component of thinking skills, making individuals critical consumers of scientific knowledge and practical problem-solvers (Pizzini & Shepardson, 1991). The posing or formation of a good question by students not only activates their prior knowledge, but also helps them elaborate on their knowledge. This is the heart of what doing science is all about (Dkeidek, Mamlok-Naaman, & Hofstein, 2011; Schmidt, 1993). In addition, researchers have indicated that different kinds of problems to be coped with would
direct the learning process and influence the learning performance of students (Scardamalia & Bereiter, 1992; Sockalingam & Schmidt, 2011). Therefore, in an inquiry-based science curriculum supported by the knowledge building pedagogy, it is important for students to cultivate the abilities of exploring, confirming, or conducting procedures in inquiry activities.

Question type has been extensively studied in questioning research. Since different kinds of questions can challenge and stimulate the mind to different extents, questions can be classified according to the level of thought required to answer them. Furthermore, questions can even direct the learning process to different extents (Chin & Osborne, 2008; Scardamalia & Bereiter, 1992; Watts, Gould, & Alsop, 1997). Scardamalia and Bereiter (1992) found that a lack of domain-specific prior knowledge may influence the kinds of questions that students ask. Their study defined three question types: basic information questions, uneducated guess questions, and wonderment questions. The difference between “basic information” and “wonderment” questions depends on students’ familiarity with the topic. Graesser, Person, and Huber (1992) developed a taxonomy of questions according to cognitive science including eighteen types, while Pizzini and Shepardson (1991) suggested three categories of questions: input, processing and output, using cognitive levels as a criterion. Yet another perspective on classifying students’ questions was offered by Watts, Gould, and Alsop (1997), who described three categories of students’ questions in the process of conceptual change: consolidation questions, exploration questions, and elaboration questions. Along with their progress in terms of the types of question posed, students will also improve their questioning ability. In the beginning, they just attempt to confirm explanations, then seek to expand their knowledge, and in the end they can examine claims and counterclaims or reconcile different understandings. However, while more than half of the “raw” questions students ask do not at first lend themselves to practical investigations (Symington, 1980; Roth & Roychoudhury, 1993), it is workable to translate such questions into investigable ones with help from the teacher. Moreover, students who have experience of asking questions through the inquiry approach have been found to significantly outperform others with regard to their ability to ask more and better questions (Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005).

**Novice and experienced learners**

Mobile and ubiquitous learning, as an innovative learning strategy, seems to be a promising learning approach to support situated learning with peer communications, however, these new learning scenarios might be too complex for the students because of the requirement of integral skills application ability and sufficient prior knowledge (Hwang, Shi, & Chu, 2011). Some previous studies have pointed out that domain-specific prior knowledge might be a factor that limits the quality of questions at the beginning stage of problem-based learning (Miyake & Norman, 1979). On the other hand, researchers have indicated that students who have experience of asking questions through the inquiry approach significantly outperform those who have no such experience (Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005). Experience, according to the literature (Ertmer & Newby, 1996; Spires & Donley, 1998), is generally regarded as one of the most crucial elements for successful and efficient learning. Researchers have indicated that naïve and expert learners display many different learning characteristics, no matter whether they learn in traditional or digital instructional environments (Chen, Fan, & Macredie, 2006; Williams & Noyes, 2007). Several studies have reported that experienced students often perform better than novice students with regard to both learning processes and learning results, implying the necessity of offering novice students proper supports for fostering the required abilities (Chi, Glaser, & Farr, 1988; Artino, 2008). Nevertheless, few quantitative studies have been conducted to investigate the impacts of learning experience or prior knowledge on students’ learning performance in u-learning, not to mention the analysis of higher order thinking behaviors, such as problem solving or question raising (Simmons & Lunetta, 1993), which need proper learning supports as well as sufficient practice. Problem solving strategies acquired from experienced learners could be useful to naïve learners. Furthermore, from the perspective of cognitive psychology, cognitive activities are highly structured; that is, specific hierarchies or constructs for describing the relationships between knowledge items or concepts exist, showing the importance of learning design. This also implies the need to guide students to acquire basic or prior knowledge before learning advanced learning contents in designing school curricula. Moreover, it can be inferred that taking students’ learning experiences or prior knowledge into account is important for designing u-learning activities.
Collaborative learning

In the past decades, collaborative learning has been seen as an effective teaching method and learning strategy (Jacob, 1999; Johnson & Johnson, 1999; McInerney & Robert, 2004). A large number of studies have shown the benefits of collaborative learning in terms of improving learners’ cognitive achievement, learning motivation, and peer relationships (Schoor & Bannert, 2011). Meanwhile, various collaborative learning techniques and instructional skills have been developed and applied in different learning situations, including Jigsaw II (Sahin, 2010; Slavin, 1986) and Learning Together (Johnson & Johnson, 1999), for fostering learning and elaborating teaching. After many years of evolution, science inquiry has been defined as a process of identifying and posing questions, searching for information, designing and carrying out scientific investigations, analyzing data and making conclusions, creating artifacts, and sharing and communicating findings (Krajcik et al., 1998; NRC, 1996; Sun & Looi, 2013). This learner-centered learning method emphasizes the application of classroom-learned knowledge to realistic scenes as well as the importance of concern for personal living surroundings and exploring novel and meaningful questions for practical use or thorough understanding.

Following the recent rapid advancements in information technology, computer-supported collaborative learning (CSCL) has become a potential direction for scaffolding students’ critical thinking and problem solving (Salomon, Perkins, & Globerson, 1991; Jonassen, 1996). Many researchers have identified the potential of using computer systems to support collaborative learning activities (Fan & Lê, 2011; Mason & Watts, 2012; Morris, 2008; Vonderwell, Liang, & Alderman, 2007; Xie & Bradshaw, 2008). For example, several studies have employed the computer-supported collaborative learning (CSCL) approach to conducting PBL activities (Resta & Laferrière, 2007; Rientes et al., 2012), in which students can develop their collaborative learning skills through the activities of problem exploration, peer discussion, and problem solving in the process of PBL with the assistance of technological tools (Lu, Lajoie & Wiseman, 2010). Contrast with the advantages of applying technology, there are still four challenges faced in the implementation of contemporary science CSCL environments, namely (1) Most applications do not seem to be robust enough to support social interaction, quick feedback and evaluation across distances and at different times; (2) Few applications are available to support synchronous collaboration; (3) With their flexibility limitations, most of these environments are not appropriate for a wide range of activities in different science subject areas; and (4) Most systems are not comprehensive enough to combine inquiry, modeling and collaborative learning approaches to facilitate students’ development of critical learning skills in science (Dimitracopoulou et al., 1999; Sun & Looi, 2013).

With the development of technology, computer-supported environments are no longer limited to indoors. Furthermore, the use of mobile technologies has also become more popular for collaborative science inquiry because of the advantages of portability and information retrieval which can occur at any time and in any place (Hung, Hwang, Lee, & Wu, 2011; Hwang, Shih, & Chu, 2011; Shih, Chuang, & Hwang, 2010; Vogel, Spikol, Kurti, & Milrad, 2010). Students can not only discuss questions, exchange opinions, and share information with peers or instructors anytime and anywhere, but they can also acquire their learning experiences from real-world learning tasks (Looi et al., 2009; Wong, 2012). Researchers have indicated that, in a well-designed curriculum with effective technology supports and appropriate contextual environment settings, students are able to develop both competencies of scientific literacy and problem-solving skills during the learning process (Zhang et al., 2010, Hwang, Shi, & Chu, 2011). With the assistance of mobile and wireless communication technologies, learners can discuss with their peers in different places simultaneously and search for relevant data or useful information on the web to solve problems.

Development of a ubiquitous problem-based learning system (UPBLS)

In this study, a ubiquitous problem-based learning system (UPBLS) was developed based on the assumption that most novice students start learning by asking intuitively interesting questions, which may not at first seem to be realistic or scientifically relevant. It is expected that, with the help of mobile and wireless communication and sensing technologies, students are able to gain experience in the field and link what they have observed in the real world with what they have learned from the textbooks. Via accumulating experience of peer discussion, data collection, idea sharing, and reflective diary writing in the field trip, students are able to successfully refine their
problems by adding scientific elements, until they finally become quality scientific problems. In the meantime, it is expected that the students’ scientific inquiry competence can be improved as well (see Figure 1).

Based on the UPBL model (see Figure 1), the UPBLS is developed to optimize the collaboration of science learning communities. The central working area in UPBLS is composed of group tasks. Learners are able to edit their notes, diaries, and reports. Besides the central group task area, UPBLS provides three other mechanisms: an On-line discussion forum, an E-library, and a web-based visualization tool (called the Green Lab). The e-library provides rich and relevant data that stimulate students to learn and think more during the field trip. The green lab enables students to summarize and present their findings in the field, which helps them think from different aspects. Moreover, the questioning rubrics guide and encourage students to propose quality questions based on what they have observed and found during the learning activity.

![A triangular model of the UPBLS design](image)

With these three functions, students are guided to accomplish their group tasks and to refine their research questions step by step. Combined with the central group task area and three other mechanisms, students can get access to the UPBL system not only via the computers in school after performing the inquiry activities, but also via smartphones during their field observations. The on-line discussion forum (see Figure 2) helps students reflect on, clarify, stimulate, and monitor their inquiries. Everyone who participates in the inquiry activities can respond to others’ subjects or propose a new subject. The E-library (see Figure 3) contains an ecology database designed to help the learners to refine their questions. Students can search for information when describing or recording their findings of the organisms in their natural environment. They can also access the detailed information by means of QR codes through smartphones if they are interested in a particular creature.

Furthermore, measurement statistics are provided in the Green Lab (Vogel, Kurti, Milrad, & Kerren, 2011) in UPBLS to present the collected data (see Figure 4). The need for web-based visualization tools in this area indicates the importance of allowing learners in an interactive manner to explore, analyze and reflect on different representations of environmental data (Vogel, 2011). For instance, salinity, pH value, dissolved oxygen in water, turbidity and temperature were under investigation and were recorded in the database. Statistics will change along with the data recorded. Learners can search for the information they need in the Green Lab and utilize the acquired knowledge to formulate scientific questions or to work on their reports. The Green Lab was developed to update and share all ecological observation information. Moreover, it enables the visualization of different types of geo-tagged content and sensor data collected using mobile devices.
Figure 2. The interface of group tasks and on-line discussion

Figure 3. The interface of the E-library
Research design

Based on the UPBLS triangle model, the research combined both the collaborative learning and ubiquitous learning approaches. Therefore, collaboration with the advantage of ubiquitous technology to propose and refine problems is the main concept embedded in the research inquiry activities. Learners can not only carry out ecological observations, collect data and record information in the inquiry activities, but can also keep on-line learning diaries, carry out on-line discussion and apply measurement statistics in group reports after the inquiry activities. Through this learning approach, the learners can first propose intuitive problems that they are interested in after gathering and receiving all ecological information and then refine them gradually to become workable, such that finally, a scientific problem can be proposed. To assess the learners’ progress in terms of their questioning ability and inquiry competence, scoring rubrics for questioning ability and scientific inquiry literacy assessment, Computerized Scientific Inquiry Literacy Assessment (CSILA), were applied.

Participants

A total of 43 sixth and fifth graders participated in the UPBL program, of whom 25 sixth graders were defined as experienced students who had gone through six months of inquiry activities before the experiment (Hung, Hwang, Lee, & Wu, 2011). The other 18 5th grade pupils, who had never joined any inquiry activity before the experiment, were categorized as novices. That is, the students were categorized into the two groups mainly based on their prior inquiry activity experience instead of their scientific inquiry literacy or learning performance.

Procedure and learning scenarios

Three UPBL field observation activities were arranged between November 2011 and May 2012, as shown in Table 1. CSILA was administered three times, once before the inquiry activities and twice after the activities, to assess the
The performance of the students’ inquiry ability via evaluating their questioning performance. The reasons for conducting the test twice after the activity were to investigate the correlation patterns between the students’ inquiry performance and questioning ability and to illustrate their progress trend. Following the first CSILA, the anchored instructions provided the students with an introduction to the wetland. Furthermore, instruction was given to help the students become familiar with the operations of the smartphones and scientific instruments. In the following three inquiry activities, the students were supported by UPBL to raise questions, gather data, discuss with team members, revise questions and finally share the outcomes with other teams. During the field observations, each participant was equipped with a smartphone, which was used to interact with the learning system as well as to gather information for accomplishing the PBL tasks.

<table>
<thead>
<tr>
<th>Date</th>
<th>Stages</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011/11</td>
<td>Test 1</td>
<td>1st CSILA administered</td>
</tr>
<tr>
<td>2011/12</td>
<td>Anchored Instructions</td>
<td>a. Introduction of Sihcao Wetland</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. Application of smartphone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. Operation of instruments</td>
</tr>
<tr>
<td>2011/12</td>
<td>Inquiry Activity 1</td>
<td>a. First trip to three different locations of Sihcao Wetland to investigate the characteristics of the water</td>
</tr>
<tr>
<td>2012/1</td>
<td></td>
<td>b. Sharing initial thoughts about the inquiry problems</td>
</tr>
<tr>
<td>2012/1</td>
<td>Inquiry Activity 2</td>
<td>a. Second trip to the Sihcao Wetland</td>
</tr>
<tr>
<td>2012/3</td>
<td></td>
<td>b. Sharing revised inquiry plans and measurement results</td>
</tr>
<tr>
<td>2012/3</td>
<td>Inquiry Activity 3</td>
<td>a. Third trip to Sihcao Wetland to execute group plans</td>
</tr>
<tr>
<td>2012/4</td>
<td></td>
<td>b. Sharing their preliminary results and revised plans</td>
</tr>
<tr>
<td>2012/4</td>
<td>Oral presentations</td>
<td>Face to face discussions</td>
</tr>
<tr>
<td>2012/4</td>
<td>Test 2</td>
<td>2nd CSILA administered</td>
</tr>
<tr>
<td>2012/5</td>
<td>Test 3</td>
<td>3rd CSILA administered</td>
</tr>
</tbody>
</table>

The students were supported by the ubiquitous learning system to complete the three stages of the inquiry tasks, as shown in Figure 5. To accomplish the inquiry tasks and raise specific scientific problems successfully, they were asked to collect two categories of data via making observations and measurement: (1) environmental data, such as the quality of water and air; and (2) information about the creatures in the ecological area. In each of the three stages, the students were scheduled to collect data. Moreover, the data to be collected in one stage depended on the findings and group discussion results of the previous stage.

![Figure 5. Learning scenarios of the ubiquitous problem-based learning activity](image)

In the first stage, the students were guided in the field to make observations and measurements to collect environmental data based on the worksheets displayed on the mobile devices, which aimed to intuitively situate them.
in question-raising scenarios. In the second stage, the students were asked to carry out on-line discussions and self-reflections on the group task platform after the field trips. Via information sharing and discussion, each group of students refined the questions they raised.

In the final stage, the students were asked to review the data they had measured and collected, including the statistical results, to find out the required scientific evidence for solving their scientific questions. After each field trip, individual students were asked to complete a learning diary on the website based on what they had observed and learned.

During the inquiry activities, the students could log into the website to complete the learning sheets, as shown in Figure 6. In addition to the group members, the data collected by the students were also shared with other groups. The students collected these data based on the working items displayed on the mobile devices as well as the questions raised by the teacher and themselves during the field trips by inputting the observed or measured data, taking notes and photos, and searching for data from the e-library. They could then share their findings with their team members.

![Interface of the inquiry learning sheets on the mobile devices](image)

**Figure 6.** Interface of the inquiry learning sheets on the mobile devices

**Measuring tools**

In this study, the Computerized Scientific Inquiry Literacy Assessment (CSILA), developed by Hung, Hwang, Lin, Hung and Wu (2010), was integrated into the field inquiry activities to investigate the students’ scientific inquiry skills and progress. The facets included in CSILA are observation, inference and experiment design, with three different item types: observation of photos, movie clips and concept mapping. Each facet was divided into three levels: basic, proficient, and advanced. Figures 7 to 10 show the sample items of CSILA.

Figure 7 is an observation facet item embedded with video clips for observing. Figure 8 is a fill-in-the-concept-map facet item. Students can drag the right side answers to fill in the blanks according to the instructions. Figure 9 is a scientific inference facet item. Students need to draw the inferences according to the description, illustration and...
representation of the item. Figure 10 shows a sample item of the experiment design facet, including the design, illustration, and process of the experiment.

As shown in Table 2, the total number of CSILA items is 56, including 18 items for Level 1, 21 items for Level 2, and 17 items for Level 3. The average difficulty (p value) of the items is 0.62 and the Cronbach’s α of the measure is 0.71.

<table>
<thead>
<tr>
<th>Content</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>18</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Inference</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Experiment Design</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>21</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 2. CSILA item distribution specification

Figure 7. Sample item of the observation facet

Figure 8. Sample item of the fill-in-the-concept-map facet
If you want to illustrate the proportion of mouthbreeder in fish pond, which illustration is preferable?

Figure 9. Sample item of the scientific inference facet

【問題三】小明將實驗結果繪成下面折線圖。圖中橫軸、縱軸分別代表什麼？

The line chart represents the result of an experiment, do you know what horizontal axis and vertical axis mean?

Figure 10. Sample item of the experiment design facet
On the other hand, the scoring rubrics in Table 3 were developed for assessing the students’ questioning ability. Scholars have pointed out a number of pedagogic advantages of using grading rubrics, such as improving students’ attention and performance in learning activities (Creswell, 2009; Neuman, 2004; Fan & Lê, 2011; Solan, & Linardopoulo, 2011). In inquiry activities, students engage socially in interactive activities, including information sharing, question posing, and discussion. Questions embedded in the conversations help learners co-construct knowledge (Chin &Brown, 2000; Chin, 2004). Therefore, it is important to encourage students to pose questions and respond to other groups’ questions by taking the questioning issue into account when developing assessment rubrics (Chang, Wu, Weng, & Sung, 2012). Correcting or refining questions based on peers’ feedback is another important issue for helping students make reflections and improvements (Yu, Liu, & Chan, 2005). Accordingly, the rubrics are divided into four facets: Autonomous question posing, Assistance for others’ question posing, Autonomous question correcting, and Assistance for others’ question correcting. The two facets “Autonomous question posing” and

<table>
<thead>
<tr>
<th>Categories</th>
<th>Score</th>
<th>Content and Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>positivlearninginteraction</td>
<td>1</td>
<td>Posing questions that can promote learning, such as: strategies for collaboration</td>
</tr>
<tr>
<td>factual question</td>
<td>1</td>
<td>Posing questions that are based on prior knowledge or observation</td>
</tr>
<tr>
<td>procedural question</td>
<td>2</td>
<td>Posing questions about scientific experimental sequence</td>
</tr>
<tr>
<td>Science concept-oriented</td>
<td>3</td>
<td>Posing questions that are based on scientific concepts</td>
</tr>
<tr>
<td>question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>positivelearninginteraction</td>
<td>1</td>
<td>Posing questions that can promote learning, such as: strategies for collaboration</td>
</tr>
<tr>
<td>factual question</td>
<td>1</td>
<td>Posing questions that are based on prior knowledge or observation</td>
</tr>
<tr>
<td>procedural question</td>
<td>2</td>
<td>Posing questions about scientific experimental sequence</td>
</tr>
<tr>
<td>Science concept-oriented</td>
<td>3</td>
<td>Posing questions that are based on scientific concepts</td>
</tr>
<tr>
<td>question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>accuracy/elaboration</td>
<td>3</td>
<td>Posing questions or providing information that can help focus the learning content</td>
</tr>
<tr>
<td>promotion/continuity</td>
<td>3</td>
<td>Posing questions or providing information that can help the group elevate or extend the inquiry problem</td>
</tr>
<tr>
<td>promotion/continuity</td>
<td>3</td>
<td>Posing questions or providing information that can help focus the learning content</td>
</tr>
<tr>
<td>promotion/continuity</td>
<td>3</td>
<td>Posing questions or providing information that can help focus the learning content</td>
</tr>
</tbody>
</table>

Table 3. Scoring rubrics for questioning ability
“Assistance for others’ question posing” have the same categories, as do the facets “Autonomous question correcting” and “Assistance for others’ question correcting.” Moreover, three questioning levels (i.e., factual, procedural, and science concept-oriented) were designed by referring to those proposed by Chin, Brown and Bruce (2002) and Allmond and Makar (2010).

Results

Reliability and validity of the scoring rubrics

Three raters were invited to assess the students’ interactions in the learning system according to the rubrics defined in Table 3. The reliability of the rubrics is particularly high (r = .92) based on the ratings given by the three raters. On the other hand, it is found that questioning ability and inquiry ability show similar improvement trends (see Table 4). Therefore, inquiry ability is defined as an external criterion of validity of the questioning rubrics. As we can see in Table 5, the correlation matrix shows that questioning ability and inquiry ability have a moderately significant relationship (.37, .31 and .63, p < .05) in the three evaluations. The correlation coefficient suggests that the rubric questioning scores have appropriate validity. Furthermore, the correlation between the two abilities increases following the administration of CSILA after each activity. This increasing correlation can provide a reasonable pattern for the two abilities as validated evidence.

<table>
<thead>
<tr>
<th>Different form</th>
<th>Mean Experienced</th>
<th>Mean novice</th>
<th>Mean total</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>0.36</td>
<td>-0.32</td>
<td>.02</td>
<td>.86</td>
</tr>
<tr>
<td>I2</td>
<td>0.63</td>
<td>-0.15</td>
<td>.24</td>
<td>.72</td>
</tr>
<tr>
<td>I3</td>
<td>1.35</td>
<td>0.37</td>
<td>.86</td>
<td>.75</td>
</tr>
<tr>
<td>Q1</td>
<td>7.79</td>
<td>3.33</td>
<td>5.56</td>
<td>5.81</td>
</tr>
<tr>
<td>Q2</td>
<td>13.64</td>
<td>11.24</td>
<td>12.44</td>
<td>8.14</td>
</tr>
<tr>
<td>Q3</td>
<td>24.85</td>
<td>15.81</td>
<td>20.33</td>
<td>11.17</td>
</tr>
</tbody>
</table>

Q = questioning ability; I = inquiry ability

Table 4. Descriptive statistics of questioning and inquiry ability

<table>
<thead>
<tr>
<th></th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>Q1</th>
<th>Q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>.53**</td>
<td>.52**</td>
<td>.37*</td>
<td>.43**</td>
<td>.31*</td>
</tr>
<tr>
<td>I2</td>
<td></td>
<td></td>
<td></td>
<td>.71**</td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.55**</td>
</tr>
<tr>
<td>Q1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.68**</td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q: questioning ability; I: inquiry ability (CSILA); * p < 0.05. ** p < 0.01.

Improvements in the students’ questioning ability

This study focuses not only on the development of a UPBL system, but also on proving its usefulness by investigating the progress of students’ questioning abilities. To present this progress, the Hierarchical Linear Model (HLM) was used. HLM has the advantage of analyzing longitudinal data retrieved many times. Therefore, HLM was used to analyze the students’ scores of questioning ability from three different time points. Differences in intercepts and growth rates between the experienced 6th graders and the novice 5th graders were tested. Table 6 shows the contrast of coefficients estimated by the unconditional models of HLM. The unconditional model results indicate that the average growth rate β10 of all participants is around 7.64 (p < .01). This suggests that both groups demonstrated substantial growth in their questioning abilities, and UPBL is significantly helpful for developing students’ questioning performance. Furthermore, in order to clarify the difference in growth rate between these two groups, a group model of HLM was applied.

In the group model, the experienced 6th graders and novice 5th graders were coded as 1 and 0 respectively. Table 7 shows the contrast of coefficients estimated by the HLM group model. According to these results, the initial
difference $\beta_{01}$ is around 3.00 ($p < .05$) and the growth rate $\beta_{11}$ is around 2.29 ($p < .05$). This result suggests that the questioning ability of the experienced 6th graders was not only significantly better than that of the novice 5th graders at the initial stage, but that they also had higher growth rates throughout the whole learning process. Figure 11 displays the questioning ability progress slope contrast of the two groups. As we can see, it is obvious that the questioning ability growth rate of the experienced 6th graders was better than that of the novice 5th graders. In other words, combining the results of both the unconditional and conditioned HLM models from Tables 6 and 7, we can conclude that all students benefited from UPBL, but that the experienced 6th graders had greater improvement in their questioning abilities than the novice 5th graders.

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coe.</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $\beta_{00}$</td>
<td>5.38</td>
<td>0.69</td>
<td>7.78</td>
<td>42</td>
<td>0.00</td>
</tr>
<tr>
<td>Slope $\beta_{10}$</td>
<td>7.64</td>
<td>0.62</td>
<td>12.38</td>
<td>42</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coe.</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $\beta_{00}$</td>
<td>3.64</td>
<td>0.78</td>
<td>4.67</td>
<td>41</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta_{01}$</td>
<td>3.00</td>
<td>1.25</td>
<td>2.41</td>
<td>41</td>
<td>0.02</td>
</tr>
<tr>
<td>Slope $\beta_{10}$</td>
<td>6.31</td>
<td>0.69</td>
<td>9.10</td>
<td>41</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>2.29</td>
<td>1.13</td>
<td>2.04</td>
<td>41</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Discussion and conclusions**

Promoting the questioning ability of students has been recognized as being an important and challenging educational objective (Chin, Brown, & Bruce, 2002; Scardamalia, 2002; Tan & Seah, 2011). In this study, a ubiquitous problem-based learning system, UPBLS, is proposed for conducting in-field inquiry learning activities by providing learning guidance, an online discussion forum, an e-library and a green lab. The experimental results show that the rubrics used to assess questioning ability had reasonable reliability and validity; moreover, the UPBL approach was helpful to both experienced and novice students in improving their questioning ability. This finding is quite different from those reported by previous u-learning studies in which the technology-enhanced learning only benefited the experienced students (Chi, Glaser, & Farr, 1988; Artino, 2008). It is predictable that experienced students perform
better than novice students in the beginning of the inquiry activities; however, the progress of both groups is in fact a more important and essential issue to be investigated and verified. The finding of this study (i.e., both experienced and novice students were benefited) verifies that the UPBL approach can be applied to different levels of students. This also implies that, in a properly designed inquiry activity, if the students have enough time to become familiar with the learning system, their questioning ability can be gradually improved.

On the other hand, another finding that the experienced students had greater improvement than the novices in this study also implies that providing additional supports for novice students is needed. Such supports could be specially designed interfaces, system functions, learning guidance, feedback, or supplementary materials. As most studies related to adaptive or personalized learning mainly provide personal supports based on students’ knowledge levels, preferences or learning styles (Chiou et al., 2010; Hwang, Sung, Hung, Huang, & Tsai, 2012), it is necessary to develop adaptive learning models by taking learning experience into account.

From the ecology-investigating activity, it was also found that UPBLS functions as both a learning tool and a stimulus for question raising and peer interactions. In addition to the e-library which serves as the database from where students can obtain the required information, the album and representation of reports shared with all participants can also inspire students to come up with ideas. Besides learning functions, on-line discussion makes it convenient for students to have a chance to ask questions and respond to each other. With the aid of the UPBLS, students can not only initiate their intuitive questions, refine them into workable ones, and then finally shape up their scientific research plans, but also develop their collaborative learning competence from self-centeredness, meaningless discussion, and irresponsibility to positive interdependence, promotional interaction and individual accountability.

In addition to the ubiquitous learning platform, the questioning rubrics also played an important role in the scientific inquiry, as indicated by many researchers (Chin, Brown, & Bruce, 2002; Millar & Osborne, 1998; Scardamalia, 2002; Shodell, 1995; Tan & Seah, 2011). It not only was used to evaluate the questioning performance of the students, but also provided clear criteria and objectives to guide them to propose quality questions. Based on significantly high scorer reliability and moderate validity gained by correlating CSILA scores with the questioning rubrics score, the effective verification of the experimental results is valid.

There are, however, some limitations to the present study. First, the sample size was not large enough to support further analysis, such as comparing the learning performances of the students with different learning styles, genders and knowledge levels. Second, to apply the platform to other learning activities, the content of the e-library might require significant modifications. Furthermore, the conclusions cannot be generalized to other applications with participants of different ages.

To sum up, from the experimental results, it is concluded that the use of mobile/ubiquitous technologies in the field trip with the problem-based learning approach has provided effective supports and encouragement for improving the students’ inquiry and questioning abilities. Moreover, the proposed approach can benefit both novice and experienced students. Currently, we are planning to assess other important competences by conducting more in-field activities with UPBLS, including collaboration abilities, inquiry abilities and problem-solving abilities. Moreover, we also plan to upgrade UPBLS by implementing an automatic scoring function to provide immediate feedback to individual students.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 99-2511-S-011-011-MY3 and NSC 101-2511-S-011 -005 -MY3.

References


Identifying Computer-Supported Collaborative Learning (CSCL) Research in Selected Journals Published from 2003 to 2012: A Content Analysis of Research Topics and Issues

Lanqin Zheng, Ronghuai Huang, and Junhui Yu

1School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, 100875, China // 2Collaborative & Innovative Center for Educational Technology, Beijing Normal University, Beijing, 100875, China // bnuzhenglq@bnu.edu.cn // huangrh@bnu.edu.cn // yujunhuiibnu@gmail.com

(Submitted July 28, 2013; Revised February 3, 2014; Accepted February 18, 2014)

ABSTRACT

This study aims to identify the emerging research trends in the field of computer-supported collaborative learning (CSCL) so as to provide insights for researchers and educators into research topics and issues for further exploration. This paper analyzed the research topics, methods and technology adoption of CSCL from 2003 to 2012. A total of 706 articles from 9 leading SSCI journals were selected and analyzed through the lens of research topics, research design, research methods, data sources, data analysis methods, research settings, research sample groups, research learning domain, types of collaborative learning, and technology for supporting collaborative learning. The results indicated that technology support of collaborative learning, interaction pattern and analysis, CSCL evaluation, and CSCL practice or applications in naturalistic teaching settings and community environment were the four major research topics. Qualitative research approach and descriptive research method were found in most of CSCL publications. Furthermore, it was found that the research topics significantly correlated with the research design and research methods. In the future, researchers and educators should pay more attention to technical affordance and facilitation in collaborative learning.

Keywords

Research trends, Research topics, Research methods, Computer-supported collaborative learning, Content analysis

Introduction

Computer-supported collaborative learning (CSCL) has become more and more popular and numerous studies related to CSCL have been reported. As an area of the learning sciences, the main concern of CSCL focuses on how people can learn together with the aid of computers (Stahl, Koschmann, & Suthers, 2006). In this area, a wide range of topics have been explored, such as CSCL system design, argumentative activities, CSCL scripts, interaction analysis etc. Furthermore, various kinds of research methods are adopted to understand the complex interactional phenomena and transform educational practices. Therefore, it is very significant and valuable to review and examine the CSCL trend in research topics and research issues so as to provide good references for researchers in this field.

In recent years, some reviews on CSCL researches have been conducted. These review had various kinds of foci, including review of the systems that supported the management of collaborative learning interaction (Soller et al., 2005), meta-analysis of research methods in CSCL field (Resta & Laferrière, 2007; Jeong & Hmelo-Silver, 2011), identifying the pitfalls for social interaction in CSCL environments (Kreijns et al., 2003), analysis of research foci on argumentation-based computer supported collaborative learning (Noroozi et al., 2012), review of the research topics in the International Journal of Computer-supported Collaborative Learning from 2006 to 2011 (Lonchamp, 2012), and review of the nine experimental studies on the impact of mobile technologies in supporting collaborative learning (Hsu & Ching, 2013). However, very few studies have thoroughly analyzed the research topics, research methodology, and technology adoption in this field over the past decade.

The present study provides insight into the research trends in the field of CSCL by systematically analyzing articles published in the nine selected SSCI journals on CSCL from 2003 to 2012. The nine selected journals were Computers & Education (C&E), International Journal of Computer-Supported Collaborative Learning (ijCSCL), Journal of Computer Assisted Learning (JCAL), Instructional Science (I&S), Educational Technology Research & Development (ETR&D), British Journal of Educational Technology (BJET), Australasian Journal of Educational Technology (AJET), Interactive Learning Environments (ILE), and Educational Technology & Society (ET&S). The purpose of this study has been twofold. First, we investigate the distribution of research topics, research
methodology, and technology adoption among the articles published in the nine leading journals during 2003 to 2012. Second, we explored the emerging trends after reviewing the nine journal publications. Therefore, the four research questions addressed in this study are as follows:

• What research topics related to CSCL were published in these selected journals from 2003 to 2012? And what were the topic variations between the first five years (2003-2007) and the second five years (2008-2012)?

• What research methodology related to CSCL was selected in these selected journals from 2003 to 2012? And what were the methodology variations between the first five years (2003-2007) and the second five years (2008-2012)?

• What types of collaborating learning and technologies for supporting collaborative learning were adopted in these selected journals from 2003 to 2012? And what were the collaborating learning type and technology adoption variations between the first five years (2003-2007) and the second five years (2008-2012)?

• Is there any significant association between research design, research methods and research topics?

Method

Materials

This study selected papers relevant to computer-supported collaborative learning in the SSCI database from 2003 to 2012. The journals with the top 21% of impact factor in the field of education were selected as the literature source. Furthermore, nine journals closely related to CSCL were identified, namely Computers & Education (C&E), International Journal of Computer-Supported Collaborative Learning (ijCSCL), Journal of Computer Assisted Learning (JCAL), Instructional Science (I&S), Educational Technology Research & Development (ETR&D), British Journal of Educational Technology (BJET), Australasian Journal of Educational Technology (AJET), Interactive Learning Environments (ILE), and Educational Technology & Society (ET&S). The impact factors in 2012 released by the Institute for Scientific Information (ISI) Journal Citation Reports are shown in Table 1.

<table>
<thead>
<tr>
<th>Journal title</th>
<th>Impact factor in 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computers &amp; Education (C&amp;E)</td>
<td>2.775</td>
</tr>
<tr>
<td>International Journal of Computer-Supported Collaborative Learning (ijCSCL)</td>
<td>1.717</td>
</tr>
<tr>
<td>Journal of Computer Assisted Learning (JCAL)</td>
<td>1.632</td>
</tr>
<tr>
<td>Instructional Science (I&amp;S)</td>
<td>1.568</td>
</tr>
<tr>
<td>Educational Technology Research &amp; Development (ETR&amp;D)</td>
<td>1.522</td>
</tr>
<tr>
<td>The Australasian Journal of Educational Technology (AJET)</td>
<td>1.363</td>
</tr>
<tr>
<td>British Journal of Educational Technology (BJET)</td>
<td>1.313</td>
</tr>
<tr>
<td>Interactive Learning Environments (ILE)</td>
<td>1.302</td>
</tr>
<tr>
<td>Educational Technology &amp; Society (ET&amp;S)</td>
<td>1.171</td>
</tr>
</tbody>
</table>

Review processes

There were a total of 6,099 documents published by these nine journals from 2003 to 2012. In order to exactly identify CSCL papers, the paper selection process proceeded in two stages. In the first stage, 807 papers were selected using keyword searches including collaborative learning, CSCL, collaborative, collaboration, online interaction, online discussion, and group discussion within the nine journals. In the second stage, authors read the full text of each paper and screened the papers one by one to decide whether they could be included. There were two criteria of paper selection. First, non-research articles such as “book reviews,” “editorials,” and “letters” were excluded from this study. Second, the study should apply information and communication technologies (ICT) to support collaborative learning in pairs, groups or community. Third, conceptual paper and empirical studies in CSCL research were included so as to produce a comprehensive representation of this body of knowledge. Finally, 706 articles were selected for the analysis in this study. These articles included 239 articles from C&E, 144 articles from ijCSCL, 85 articles from ET&S, 48 articles from JCAL, 44 articles from BJET, 38 articles from I&S, 38 articles from ETR&D, 35 articles from AJET, and 35 articles from ILE.
Coding scheme

In order to thoroughly analyze the research methodology in CSCL field, we examined the articles in the following sub-dimensions: (1) Research design, (2) Research methods, (3) Data sources, (4) Data analysis methods, (5) Research settings, (6) Research sample groups, (7) Research learning domains. For technology adoption, two sub-dimensions were analyzed: one was the type of collaborative learning; another was the technology in support of collaborative learning. The following section will illustrate the code schemes of research topics, research methodology, and technology adoption in detail.

Research topics

In order to analyze research topics and the evolution of topics over time, co-word analysis method (Callon et al., 1983) was used to identify the specific topics. The main feature of co-word analysis method was that it can extract themes and detect the linkages among these themes (Callon et al., 1986). This method used co-occurrence of pairs of words to identify relationships among ideas. In addition, co-occurrence frequencies were used to measure the relationship strength based on the words extracted from the abstract, keywords and the full text. Co-word analysis method did not require any manual coding of the corpus. This study mainly adopted the full text co-word method to analyze the research topics with the aid of the WordStat software. More specifically, in order to determine the research topic, we used term frequency weighted by inverse document frequency (TF.IDF) to obtain the results of thematic analysis. TF.IDF weighting can provide insights into highly representative themes in a larger collection of words (Salton & McGill, 1986). In addition, hierarchical clustering (Jardine & Sibson, 1971) and multidimensional scaling methods (Kruskal & Wish, 1978) were also used to map the data.

Research design

Research design was coded as: 1. Qualitative approach, 2. Quantitative approach, and 3. Mixed methods approach. A qualitative approach is one in which investigators use constructivist or participatory or both perspectives and employ inquiry strategies such as narratives, ethnographies, case studies, and so on to develop knowledge. A quantitative approach is one in which investigators use postpositivist claims for developing knowledge and employ inquiry strategies such as experiments, surveys, and so on. A mixed methods approach is one in which the researchers make knowledge claims based on pragmatic grounds and employs inquiry strategies of mixed method (Cresswell, 2009).

Research methods

Traditionally, CSCL research is characterized as three main methodologies: iterative design, experimental, and descriptive (Suthers, 2006). However, action research and mixed research emerged in recent years. Therefore, the present study coded the research methods as: (1) Experimental, (2) Descriptive, (3) Design-based research, (4) Action research, and (5) Mixed research. Experimental design refers to a planned intervention in the natural order of events by the researcher (Campbell & Stanley, 1963). Descriptive studies included surveys, case studies, and ethnographic investigations (Jeong & Hmelo-Silver, 2011). Design-based research referred to designs and interventions were conducted through continuous cycles of design, enactment, analysis, and redesign for designing learning environments and developing theories (Design-Based Research Collective, 2003). Action research referred to a form of self-reflective enquiry undertaken by participants in social situations in order to improve the rationality and justice of practices and the situations in which the practices are carried out (Carr & Kemmis, 1986). Mixed research referred to adopting hybrid research methodologies (Johnson & Onwuegbuzie, 2004) in CSCL investigations.

Data sources

Data sources were coded as: (1) Process data, including discussions transcripts, email messages, video records, audio records, log data, think-aloud protocols, screens records, (2) Outcome data, including test, artifacts (such as products, solutions, assignments, reports, presentations), map (such as trace diagrams, concept maps, mind maps, dialog maps, argument graphs, evidence maps), (3) Miscellaneous data, including questionnaire, interview data, notes (such as
written notes, field notes, coaching diaries, written reflections, summaries, essays, journals, observation records), and (4) Non-specified.

Data Analysis methods

Data analysis method refers to how the data will be analyzed to answer research questions or test hypothesis (Fraenkel & Wallen, 2000). Methods for analysis were coded as: (1) Qualitative analysis method, including descriptive analysis, conversational analysis and discourse analysis, interaction analysis (such as contribution analysis, uptake analysis, context analysis, etc.), grounded theory, ethnographic analysis, narrative analysis, case analysis, critical event recall, think-aloud protocol analysis, and event sequence analysis; (2) Quantitative analysis method, including content analysis, statistical analysis, social network analysis, modeling analytical method (including structural equation model, hierarchical linear model, hidden Markov model etc.), lag-sequential analysis, gaming path analysis, eye-tracking analysis, participation graphs analysis, and cluster analysis.

Research settings

Research settings refer to the contexts in which the research was mainly conducted. Research settings were coded as: (1) Lab, (2) Naturalistic teaching, (3) Distance learning in formal learning, (4) Informal learning, and (5) Non-specified. If a study was concerned with informal learning in a distance learning settings, it was coded into informal learning settings.

Research sample groups

Research sample groups were categorized into one of the following sub-categories: (1) Kindergarten, (2) Primary school, (3) Junior and Senior High School, (4) Higher education, (5) Vocational education, (6) Teachers, (7) Mixed Group, and (8) Non-specified. If a study included more than one sample group, only one major sample group was coded.

Research learning domains

Research learning domains were categorized into: (1) Natural Science (including science, mathematics, physics, chemistry, biology, geography, environment science, astronomy, and architecture), (2) Social Science (including politics, education, psychology, linguistics, art, law, literature, archaeology, and philosophy), (3) Engineering & Technological Science (including engineering, computer science, and educational technology), (4) Medical Science (including health and medicine), (5) Mixed learning domain, and (6) Non-specified.

Types of collaborating learning

Types of collaborative learning refer to the context in which collaborative learning occurs. The types were coded into four categories: (1) Online collaborative learning, (2) Face-to-face collaborative learning, (3) Face-to-face and online collaborative learning, and (4) Non-specified.

Technology for supporting collaborative learning

Technology for supporting collaborative learning included: (1) Agents, (2) Online learning environment (such as platform, email, learning activity management system, other softwares that supported collaborative learning), (3) Mobile learning environment, (4) Social Media (such as blog, wiki, podcast, Facebook, Twitter), (5) Games, (6) Virtual Reality or Augment Reality, (7) Conference System, (8) Computer Supported Collaborative Learning environment (such as argumentation tools, script tools, awareness tools, knowledge forum, group scribbles, chat tools, discussion board, interaction analysis tools, etc.), (9) Groupware (such as shared environment), (10) Interactive
White Board (IWB), (11) Multi-touch/Tangible technology, (12) Multimedia teaching softwares, and (13) Non-specified. If a study included more than one kind of technology, only one major technology was coded.

**Inter-rater reliability**

The first author and three graduates majoring in educational technology manually coded all of articles according to the aforementioned categories. At first, four raters discussed the coding criteria, and then they independently coded all of the articles. The agreement rate between coders was above 0.88, regarded as an indication of reliable results and good agreement. The four raters discussed and resolved all discrepancies.

**Results**

**Research question 1: What research topics related to CSCL were published in these selected journals from 2003 to 2012? And what were the topic variations between the first five years (2003-2007) and the second five years (2008-2012)?**

**Research topics**

In order to analyze research topics, we firstly got the higher frequency terms with the TF.IDF weights in the full text corpus, abstracts corpus and keywords corpus. Table 2 showed the 20 selected token types. As is shown in Table 2, some terms closely related to collaborative learning such as “wiki,” “script,” “message,” “argumentation,” and “code” have emerged.

<table>
<thead>
<tr>
<th>Token type</th>
<th>Full text corpus</th>
<th>Abstracts corpus</th>
<th>Keywords corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game</td>
<td>4579</td>
<td>2808.7</td>
<td>1951</td>
</tr>
<tr>
<td>Tutor</td>
<td>4531</td>
<td>2105.1</td>
<td>898</td>
</tr>
<tr>
<td>Annotation</td>
<td>2205</td>
<td>2007.1</td>
<td>709</td>
</tr>
<tr>
<td>Wiki</td>
<td>2181</td>
<td>1941.2</td>
<td>594</td>
</tr>
<tr>
<td>Script</td>
<td>2524</td>
<td>1928.7</td>
<td>487</td>
</tr>
<tr>
<td>CSCL</td>
<td>4638</td>
<td>1819.9</td>
<td>426</td>
</tr>
<tr>
<td>Team</td>
<td>6421</td>
<td>1761.9</td>
<td>429</td>
</tr>
<tr>
<td>Message</td>
<td>6362</td>
<td>1721.5</td>
<td>449</td>
</tr>
<tr>
<td>Argumentation</td>
<td>2419</td>
<td>1595.6</td>
<td>403</td>
</tr>
<tr>
<td>Online</td>
<td>11052</td>
<td>1591.5</td>
<td>520</td>
</tr>
<tr>
<td>Pupil</td>
<td>1640</td>
<td>1549.3</td>
<td>623</td>
</tr>
<tr>
<td>ICT</td>
<td>2106</td>
<td>1514.8</td>
<td>193</td>
</tr>
<tr>
<td>Phase</td>
<td>4782</td>
<td>1396.2</td>
<td>463</td>
</tr>
<tr>
<td>Teacher</td>
<td>16947</td>
<td>1359.1</td>
<td>425</td>
</tr>
<tr>
<td>Score</td>
<td>4414</td>
<td>1333.2</td>
<td>465</td>
</tr>
<tr>
<td>Condition</td>
<td>4980</td>
<td>1328.7</td>
<td>549</td>
</tr>
<tr>
<td>Mobile</td>
<td>1790</td>
<td>1292.3</td>
<td>350</td>
</tr>
<tr>
<td>Code</td>
<td>4557</td>
<td>1250.4</td>
<td>508</td>
</tr>
<tr>
<td>Representation</td>
<td>3861</td>
<td>1246.8</td>
<td>241</td>
</tr>
<tr>
<td>Instructor</td>
<td>3220</td>
<td>1231.5</td>
<td>253</td>
</tr>
</tbody>
</table>

With the aid of hierarchical clustering, the research topic analysis was elaborately conducted in the full text corpus. In this study, an average-linkage hierarchical clustering method was adopted to create clusters from a similarity matrix. However, the size of matrix for hierarchical clustering cannot be too large or too small, otherwise it would have been very difficult to draw and analyze (Ding et al., 2001; Lonchamp, 2012). We tried to test five kinds of word-word cooccurrence matrices: 500x500 matrix, 400x400 matrix, 300x300 matrix, 200x200 matrix and 120x120 matrix. The result indicated that 200x200 matrix that generated 36 clusters can better interpret the research topics. As is shown in Figure 1, different colors denote different clusters and the size of circle denotes the weights. The most obvious themes with the higher frequency words in bold italic fonts are given as follows (They are boxed in Figure 1):
RT1 = \{\textit{System}, software, web, user, interface, mobile, device, digital, service, platform, PC, document, annotation, management, search, resource\}
RT2 = \{\textit{Game}, action, player, simulation, element\}
RT3 = \{\textit{Wiki}, wikis, edit, page\}
RT4 = \{\textit{Blog}, blogging\}
RT5 = \{\textit{Affordances}, IWB, GS\}
RT6 = \{\textit{Object}, rule, component\}
RT7 = \{\textit{Agent}, constraint\}
RT8 = \{\textit{Program}, mentor\}
RT9 = \{\textit{Flow}, IMS\}
RT10 = \{\textit{Module}, blend, PBL\}
RT11 = \{\textit{Semantic}, ontology\}
RT12 = \{\textit{Message}, synchronous, asynchronous, CMC, conference, video, forum, chat, posting, conversation, thread, contribution\}
RT13 = \{\textit{Argument}, argumentation, claim, argumentative, sequence, elaboration, explanation\}
RT14 = \{\textit{Debate}, graph, diagram\}
RT15 = \{\textit{Code}, analyze, scheme, category, unit, discourse, pattern, frequency\}
RT16 = \{\textit{Segment}, KC\}
RT17 = \{\textit{Dialogue}, utterance, sentence, statement\}
RT18 = \{\textit{Write}, text, essay, read, comment\}
RT19 = \{\textit{Test}, score, post, effect, performance, measure, posttest, treatment, experiment, hypothesis, variable, correlation, difference, experimental, control, condition\}
RT20 = \{\textit{Response}, achievement, attitude, interview, survey, questionnaire, item\}
RT21 = \{\textit{Conceptual}, epistemic, artifact\}
RT22 = \{\textit{Perceive}, satisfaction, presence, distance\}
RT23 = \{\textit{CSCL}, script\}
RT24 = \{\textit{Scaffold}, metacognitive, prompt, monitor\}
RT25 = \{\textit{Behavior}, intervention, indicator\}
RT26 = \{\textit{Collective}, efficacy, belief\}
There was a big cluster that was difficult to interpret in Figure 1. However, it can be easily interpreted in the dendogram (see Figure 2.) The following eight sub-clusters that were boxed can be easily detected in Figure 2.

RT29 = {Feedback, assessment, peer, tutor, evaluation, plan, assignment, grade, instructor}
RT30 = {Community, building, network, practice, inquiry}
RT31 = {Class, classroom, student, teacher, teach, curriculum, school, ICT, science}
RT32 = {Online, participant, participation, member, team, project, space, face, virtual}
RT33 = {Pedagogic, principle, lecture, session, week}
RT34 = {Learner, cognitive, solve, task, strategy, model, instructional, strategy, solution}
RT35 = {Concept, map, expert, representation, domain}
RT36 = {Construction, negotiation, phase}

It is very obvious that RT1, RT2, RT3, RT4, RT5, RT6, RT7, RT8, RT9, RT10, and RT11 are closely located and can be merged (see Figure 1). They are related to technology and environment issues in CSCL field. RT12, RT13, RT14, RT15, RT16, RT17, and RT18 are adjacent to each other and they focus on interaction patterns and interaction analysis. RT19, RT20, RT21, and RT22 are close to each other and they deal with assessment in CSCL. From RT23 to RT36 can be categorized into CSCL practice or applications in classroom and community. Because dimension interpretation is not direct in the multidimensional scaling, four high level thematic categories are suggested as follows (They are outlined in Figure 3):

- Clusters in the upper part of the map deal with technology and environment issues in CSCL, such as media development and applications, technical affordances, collaborative learning environment and semantic technology (system, mobile, device, platform, game, blog, wiki, agent, affordance…)

- Clusters in the right part mainly deal with interaction pattern and interaction analysis in CSCL (argumentation, debate, chat, message, analyze, code, scheme…)
Clusters in the bottom of the map are closely related to assessment methods and content (test, posttest, measure, score, achievement, satisfaction, presence, perceive, conceptual, epistemic, artifact…)

Clusters in the left part of the map deal with CSCL practice or applications in classroom and community, collaborative learning support, mathematical applications (community, online, classroom, pedagogic, instructional, intervention, scaffold, feedback, mathematic…)

Figure 3. A. Technology and environment  B. Interaction pattern and analysis  C. Evaluation  D. Practice or applications in classroom and community

Table 3 shows the results of research topics analysis. “Technology in support of collaborative learning,” “Interaction pattern and analysis,” “CSCL evaluation,” and “CSCL practices or applications in natural teaching and community” are the four high level research topics that emerge. Each of them is divided into more focused themes characterized by a list of keywords.

<table>
<thead>
<tr>
<th>Research topics</th>
<th>Themes</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology in support of collaborative learning</td>
<td>System development and social media applications</td>
<td>System, software, web, interface, mobile, device, digital, platform… Game, action, player… Blog, blogging… Wiki, wikis, edit…</td>
</tr>
<tr>
<td>Technical affordances</td>
<td>Affordances, IWB, GS… Object, component, rule… Agent, constraint Program, mentor</td>
<td></td>
</tr>
<tr>
<td>Collaborative learning environment</td>
<td>IMS, flow… Module, blend, PBL…</td>
<td></td>
</tr>
<tr>
<td>Semantic technology</td>
<td>Semantic, ontology…</td>
<td></td>
</tr>
</tbody>
</table>
Interaction pattern and analysis

<table>
<thead>
<tr>
<th>Interaction pattern</th>
<th>Interaction analysis</th>
<th>CSCL evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction pattern</td>
<td>Message, chat, synchronous, asynchronous…</td>
<td>Test, posttest, measure, score, experiment…</td>
</tr>
</tbody>
</table>

### Thematic evolutions

In order to examine thematic evolution of CSCL field in past decade, word frequencies evolution based on the full text corpus was analyzed in the two consecutive periods, namely the first five years (2003-2007) and the second five years (2008-2012). We computed the deviations and the significance of word frequencies between the first five years (Freq₁, column) and the second five years (Freq₂, column) to analyze the trend of thematic evolution. P-values denote the period-to-period differences, as is shown in Table 4. The results showed that two themes became more and more popular with significant increasing deviations:

- Technical affordance and application of social media, with the emerged terms “recommender,” “tagging,” “visualization,” “wiki,” and “blog.”
- Facilitation of CSCL, with the terms “mediate,” “moderation,” “scaffold,” and “feedback.”

Meanwhile, the results also suggested that the popularity of three themes dropped with significant decreasing deviations:

- CSCL environment, with the terms “WEBCT,” “KF,” “CSILE,” and “videoconference.”
- Assessment of CSCL, with the terms “assessment,” and “evaluation.”
- Interaction pattern and interaction analysis, with the terms “counterargument,” “argumentative,” “intersubjectivity,” “debate,” “message,” “chat,” and “SNA.”

<table>
<thead>
<tr>
<th>Words</th>
<th>Freq₁</th>
<th>Freq₂</th>
<th>Dev.</th>
<th>Words</th>
<th>Freq₁</th>
<th>Freq₂</th>
<th>Dev.</th>
<th>p(2-tails)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mediate</td>
<td>1506</td>
<td>49.2</td>
<td>2959.6%</td>
<td>counterargument</td>
<td>46</td>
<td>436.0</td>
<td>-89.4%</td>
<td>.000</td>
</tr>
<tr>
<td>Blog</td>
<td>688</td>
<td>30.5</td>
<td>2157.9%</td>
<td>Media</td>
<td>170</td>
<td>1558.7</td>
<td>-89.1%</td>
<td>.000</td>
</tr>
<tr>
<td>Wikis</td>
<td>737</td>
<td>49.2</td>
<td>1397.3%</td>
<td>WEBCT</td>
<td>64</td>
<td>396.1</td>
<td>-83.8%</td>
<td>.000</td>
</tr>
<tr>
<td>Avatar</td>
<td>608</td>
<td>49.2</td>
<td>1135.2%</td>
<td>KF</td>
<td>30</td>
<td>152.4</td>
<td>-80.3%</td>
<td>.000</td>
</tr>
</tbody>
</table>
Research question 2: What research methodologies related to CSCL were selected in these selected journals from 2003 to 2012? And what were the methodology variations between the first five years (2003-2007) and the second five years (2008-2012)?

Research design and research method

Table 5 shows the descriptive data for the results of research design and research method. With regard to the research design, “Qualitative approach” research was conducted most and “Quantitative approach” research was utilized least in the two consecutive periods. Moreover, there is a research trend of decrease in “Qualitative approach” and increase in “Quantitative approach.” However, the significant difference between the initial five years (2003-2007) and the later five years (2008-2013) was only found in “Quantitative research” ($\chi^2 = 11.235, p < 0.01$). With respect to the research method, “Descriptive Research method” was found in most of CSCL publications in the two periods. “Action Research method” was the least adopted. However, there were no significant difference between the first five years (2003-2007) and the second five years (2008-2012) in “Descriptive research method” ($\chi^2 = 0.174, p > 0.05$), “Experimental Design method” ($\chi^2 = 0.417, p > 0.05$), “Design-based Research method” ($\chi^2 = 0.592, p > 0.05$), “Action Research method” ($\chi^2 = 0.112, p > 0.05$) and “Mixed Research method” ($\chi^2 = 0.001, p > 0.05$).

<table>
<thead>
<tr>
<th>Research design</th>
<th>Total n (%)</th>
<th>2003-2007 n (%)</th>
<th>2008-2012 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative approach</td>
<td>318(45.0)</td>
<td>105(48.8)</td>
<td>213(43.4)</td>
</tr>
<tr>
<td>Quantitative approach</td>
<td>165(23.4)</td>
<td>33(15.3)</td>
<td>132(26.9)</td>
</tr>
<tr>
<td>Mixed methods approach</td>
<td>223(31.6)</td>
<td>77(35.9)</td>
<td>146(29.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research method</th>
<th>Total n (%)</th>
<th>2003-2007 n (%)</th>
<th>2008-2012 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Research method</td>
<td>392(55.5)</td>
<td>122(56.7)</td>
<td>270(55)</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>239(33.9)</td>
<td>69(32.1)</td>
<td>170(34.6)</td>
</tr>
<tr>
<td>Design-based Research</td>
<td>24(3.4)</td>
<td>9(4.3)</td>
<td>15(3.1)</td>
</tr>
<tr>
<td>Action Research</td>
<td>8(1.1)</td>
<td>2(0.9)</td>
<td>6(1.2)</td>
</tr>
<tr>
<td>Mixed Research method</td>
<td>43(6.1)</td>
<td>13(6)</td>
<td>30(6.1)</td>
</tr>
</tbody>
</table>

Data sources and data analysis methods

The descriptive data for the results of data sources and data analysis methods were shown in Table 6. Researchers collected a variety of data sources in CSCL research. From 2003 to 2007, “Process Data” was utilized most. However, “Miscellaneous data” increased to 92.8% from 2008 to 2012. The total percentages exceeded 100% because some studies collected both “Process Data” and “Outcome Data.” However, the significant difference between the initial five years (2003-2007) and the later years (2008-2012) was only found in “Miscellaneous data” ($\chi^2 = 78.2, p < 0.001$).
Diverse data analysis methods were used in CSCL investigations. Among the qualitative analysis method, descriptive analysis method was the most frequently used in the two periods. Both grounded theory methods and narrative analysis methods were the least adopted. Among the quantitative analysis method, statistical analysis method was used the most. Eye-tracking analysis method, gaming path analysis, and participation graphs analysis were the least employed in the two periods. The total percentages exceeded 100% because some studies adopted both qualitative analysis method and quantitative analysis method. However, there were significant decrease in case analysis method ($\chi^2 = 6.78, p < 0.01$) and critical event recall ($\chi^2 = 9.18, p < 0.01$) between the first five years and the second five years.

**Table 6.** Descriptive data for the results of data sources and data analysis methods

<table>
<thead>
<tr>
<th>Data sources</th>
<th>Total n (%)</th>
<th>2003-2007 n (%)</th>
<th>2008-2012 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Data</td>
<td>516(73.1)</td>
<td>162(75.3)</td>
<td>354(72.1)</td>
</tr>
<tr>
<td>Outcome Data</td>
<td>267(37.8)</td>
<td>81(37.7)</td>
<td>186(37.8)</td>
</tr>
<tr>
<td>Miscellaneous data</td>
<td>600(84.9)</td>
<td>144(66.9)</td>
<td>456(92.8)</td>
</tr>
<tr>
<td>Non-specified data</td>
<td>40(5.6)</td>
<td>16(7.4)</td>
<td>24(4.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Qualitative analysis method</th>
<th>Total n (%)</th>
<th>2003-2007 n (%)</th>
<th>2008-2012 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive analysis</td>
<td>393(55.6)</td>
<td>129(60)</td>
<td>264(53.7)</td>
</tr>
<tr>
<td>Discourse analysis</td>
<td>42(5.9)</td>
<td>12(5.6)</td>
<td>30(6.1)</td>
</tr>
<tr>
<td>Grounded theory</td>
<td>2(0.3)</td>
<td>1(0.47)</td>
<td>1(0.2)</td>
</tr>
<tr>
<td>Interaction analysis</td>
<td>10(1.4)</td>
<td>3(1.4)</td>
<td>7(1.4)</td>
</tr>
<tr>
<td>Case analysis</td>
<td>145(20.5)</td>
<td>57(26.5)</td>
<td>88(17.9)</td>
</tr>
<tr>
<td>Ethnographic analysis</td>
<td>5(0.7)</td>
<td>1(0.4)</td>
<td>4(0.8)</td>
</tr>
<tr>
<td>Narrative analysis</td>
<td>2(0.3)</td>
<td>1(0.4)</td>
<td>1(0.2)</td>
</tr>
<tr>
<td>Critical event recall</td>
<td>4(0.5)</td>
<td>4(1.8)</td>
<td>0</td>
</tr>
<tr>
<td>Event sequence analysis</td>
<td>3(0.4)</td>
<td>2(0.9)</td>
<td>1(0.2)</td>
</tr>
<tr>
<td>Think-aloud protocol analysis</td>
<td>5(0.7)</td>
<td>1(0.4)</td>
<td>4(0.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantitative analysis method</th>
<th>Total n (%)</th>
<th>2003-2007 n (%)</th>
<th>2008-2012 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content analysis</td>
<td>188(26.6)</td>
<td>52(24.2)</td>
<td>136(27.7)</td>
</tr>
<tr>
<td>Model(SEM,HLM,HMM)</td>
<td>19(2.6)</td>
<td>5(2.3)</td>
<td>14(2.85)</td>
</tr>
<tr>
<td>Lag-sequential Analysis</td>
<td>6(0.8)</td>
<td>2(0.9)</td>
<td>4(0.8)</td>
</tr>
<tr>
<td>Gaming path analysis</td>
<td>1(0.1)</td>
<td>0</td>
<td>1(0.2)</td>
</tr>
<tr>
<td>Eye-tracking analysis</td>
<td>1(0.1)</td>
<td>0</td>
<td>1(0.2)</td>
</tr>
<tr>
<td>Participation graphs analysis</td>
<td>1(0.1)</td>
<td>0</td>
<td>1(0.2)</td>
</tr>
<tr>
<td>Cluster analysis</td>
<td>4(0.5)</td>
<td>0</td>
<td>4(0.8)</td>
</tr>
<tr>
<td>Social network analysis</td>
<td>17(2.4)</td>
<td>7(3.2)</td>
<td>10(2)</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>557(78.9)</td>
<td>166(77.2)</td>
<td>391(79.6)</td>
</tr>
</tbody>
</table>

**Research settings, research sample groups, and research learning domains**

Most often, CSCL research was carried out in naturalistic teaching settings, as is shown in Table 7. This was both true in the first five years (2003-2007) and the second five years (2008-2012). There were more growth in naturalistic teaching settings and lab settings between the first five years (2003-2007) and the second five years (2008-2012). But no significant differences were found in naturalistic teaching settings ($\chi^2 = 1.83, p > 0.05$) and lab settings ($\chi^2 = 2.36, p > 0.05$) between the two periods. Some research conducted in distant settings, and there was a drop between these two periods. Significant difference was also found in distant settings between these two periods ($\chi^2 = 9.23, p < 0.01$). And only a small proportion of studies were carried out in informal learning settings.

Researchers selected different sample groups in CSCL research. The “Higher Education” sample group was the most often selected and “Kindergarten” sample group was the least employed in the two periods. However, there were no significant difference in the research sample group of “Kindergarten” ($\chi^2 = 0.034, p > 0.05$), “Primary school” ($\chi^2 = 0.344, p > 0.05$), “Junior and Senior high school” ($\chi^2 = 0.236, p > 0.05$), “Higher Education” ($\chi^2 = 0.869, p > 0.05$),
“Teachers” ($\chi^2 = 0.741, p > 0.05$), “Vocational education” ($\chi^2 = 0.039, p > 0.05$), “Mixed Group” ($\chi^2 = 1.873, p > 0.05$), “Non-specified” ($\chi^2 = 2.089, p > 0.05$) between the first five years and the second five years.

Researchers selected various learning domains involved in CSCL research. From 2003 to 2007, “Engineering&Technological Science” learning domain was found in most of papers and “Medical Science” was the least selected. From 2008 to 2012, “Social Science” domain was found in most of papersand “Medical Science” was still the least selected. Furthermore, there were significant increase in “Social Science” ($\chi^2 = 3.963, p < 0.05$) and “Natural Science” ($\chi^2 = 8.392, p < 0.01$) between the first five years and the second five years. There were significant decrease in “Engineering&Technological Science” ($\chi^2 = 7.771, p < 0.01$) and “Mixed” learning domain ($\chi^2 = 11.518, p < 0.001$).

Research question 3: What kind of collaboration and technologies for supporting collaborative learning were adopted in these selected journals from 2003 to 2012? And what were the collaboration type and technology adoption variations between the first five years (2003-2007) and the second five years (2008-2012)?

Types of collaborating learning and Technology for supporting collaborative learning

Table 8 shows the descriptive data for the results of collaborating learning and technology adoption. In most of CSCL publications, “Online Collaborative Learning” was most used in the two periods. However, there were no significant difference in Online Collaborative Learning ($\chi^2 = 2.368, p > 0.05$), Face-to-Face Collaborative Learning ($\chi^2 = 1.577, p > 0.05$), Face-to-face and Online Collaborative learning ($\chi^2 = 0.006, p > 0.05$) and Non-specified ($\chi^2 = 1.086, p > 0.05$) between the first five years and the second five years.

Researchers adopted a variety of technologies that supported collaborative learning. The most frequently used technologies were “Computer Supported Collaborative Learning Environment (CSCLE)” in the two periods. Furthermore, there were significant increase in “Social Media” ($\chi^2 = 19.09, p < 0.001$), “Games” ($\chi^2 = 5.365, p < 0.05$), and “Virtual Reality or Augment Reality” ($\chi^2 = 11.16, p < 0.001$) between the first five years and the second five years.
second five years. There was significant decrease in “Computer Supported Collaborative Learning Environment” ($\chi^2 = 15.458, p < 0.001$) between these two periods.

<table>
<thead>
<tr>
<th>Types of collaborating learning</th>
<th>2003-2007</th>
<th>2008-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face-to-Face Collaborative Learning</td>
<td>107(15.2)</td>
<td>27(12.6)</td>
</tr>
<tr>
<td>Online Collaborative Learning</td>
<td>499(70.7)</td>
<td>160(74.4)</td>
</tr>
<tr>
<td>Face-to-face and Online Collaborative learning</td>
<td>80(11.3)</td>
<td>24(11.2)</td>
</tr>
<tr>
<td>Non-specified</td>
<td>20(2.8)</td>
<td>4(1.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>16(2.3)</td>
<td>3(1.4)</td>
</tr>
<tr>
<td>Online LE</td>
<td>192(27.2)</td>
<td>62(28.7)</td>
</tr>
<tr>
<td>Mobile LE</td>
<td>36(5.1)</td>
<td>11(5.1)</td>
</tr>
<tr>
<td>Social Media</td>
<td>47(6.6)</td>
<td>1(0.5)</td>
</tr>
<tr>
<td>Game</td>
<td>28(4.0)</td>
<td>3(1.4)</td>
</tr>
<tr>
<td>Virtual Reality or AR</td>
<td>31(4.4)</td>
<td>1(0.5)</td>
</tr>
<tr>
<td>Conference system</td>
<td>29(4.1)</td>
<td>11(5.1)</td>
</tr>
<tr>
<td>CSCLE</td>
<td>221(31.3)</td>
<td>89(41.4)</td>
</tr>
<tr>
<td>groupware</td>
<td>15(2.1)</td>
<td>7(3.3)</td>
</tr>
<tr>
<td>Tools developed by authors</td>
<td>57(8.0)</td>
<td>18(8.4)</td>
</tr>
<tr>
<td>IWB</td>
<td>6(0.9)</td>
<td>2(0.9)</td>
</tr>
<tr>
<td>Multi-touch/Tangible</td>
<td>2(0.3)</td>
<td>0</td>
</tr>
<tr>
<td>Non-specified</td>
<td>26(3.7)</td>
<td>7(3.3)</td>
</tr>
</tbody>
</table>

**Research question 4: Is there any significant association between research design, research methods and research topics?**

In order to analyze the relationships between research topics and research methods, Chi-Square analysis was conducted to investigate if there were significant correlations among these categories. The results showed that research topics were significantly associated with research design in the last decade ($\chi^2 = 67.52, p < 0.001$). And also research topics were also significantly associated with research methods ($\chi^2 = 141.74, p < 0.001$).

In addition, we used the adjusted residual values (AR) to examine the factors that contributed to the correlation. Only if the absolute values of AR were larger than 1.96, the factors were identified as contributing to the correlation. As shown in Table 9, the result of the adjusted residual values indicated that the topics about “Technology in support of collaborative learning” adopted the highest proportion in the “Qualitative approach” research design (residual = 3.7), followed by “Quantitative approach” research design (residual = -2.8). This kind of topic used more “Design-based method” (residual = 6.9), followed by “Experimental method” (residual = -3.4), “Mixed method” (residual = -2.5), and “Descriptive method” (residual = 2.0). The result also revealed that the research on “Interaction pattern and analysis” utilized most “Qualitative approach” research design (residual = 5.4), followed by “Quantitative approach” research design (residual = -3.1) and “Mixed methods approaches” (AR = -2.9). “Descriptive method” (residual = -6.7) was found most to use in the research on “Interaction pattern and analysis,” followed by “Experimental method” (residual = -5.9) and “Design-based research method” (residual = -2.6). With respect to the topic of “CSCL Evaluation,” “Mixed methods approaches” research design (residual = 1.7) was utilized most compared with other research design. And “Mixed” method was adopted most in CSCL evaluation research. The most prevalent research design when conducted the research on “CSCL practice or applications in natural teaching settings and community” was “Qualitative approach” research design (residual = -7.4), followed by “Quantitative approach” research design (residual = 5.5) and “Mixed methods approaches” (residual = 2.9). The most often used research method in CSCL practice studies was “Experimental method” (residual = 8.4), followed by “Descriptive method” (residual = -7.6), “Design-based research method” (residual = -3.5) and “Action research method” (residual = 2.6).
This study extended the existed analysis framework and explored the trend of the research topics, research methodology, and technology adoption in CSCL field between 2003 and 2012 based on nine SSCI indexed journals. The results from co-word analysis show that current CSCL research topics can be mainly divided into four categories: technology in support of collaborative learning, interaction pattern and interaction analysis, CSCL assessment and CSCL support or applications in classroom and community. The thematic evolutions analysis suggests that the research on “CSCL environment,” “assessment of CSCL,” and “interaction pattern and interaction analysis” gained less attention between the two five-year-periods. The researches on “technical affordance and application of social media” and “facilitation of CSCL” have significantly increased between the two consecutive periods. This result corroborated with the expectation of Clark & Mayer (2011) who think that social media such as wiki, blog, Facebook, Twitter should be integrated into the collaborative learning research. Furthermore, previous research revealed that learners typically did not engage in collaboration processes without guidance (Weinberger, Stegmann, Fischer, & Mandl, 2007). Therefore, researchers have paid more attention to facilitating collaborative learning by providing explicit scaffolding (Rienties et al., 2012) and internal and external scripts (Fischer et al., 2013). In the future, concerns related to CSCL support by scaffold, just-in-time feedback, and intervention are very predominant in this field.

Meanwhile, this study examined the trend of research methodology in detail. With respect to research design, there was a significant increase in “Quantitative approach.” More and more researchers adopted a quantitative approach to conduct CSCL research. In addition, Stahl, Koschmann, & Suthers (2006) pointed that experimental methods, descriptive methods and iterative design methods were the three traditional methods in CSCL field. The current study also revealed that “Descriptive research method” was utilized the most and “Design-based research method” was very small in CSCL publications. Also, the proportion of “Design-based research method” has decreased from 4.3% to 3.1% between the two periods. The main reason is that some challenges bound the scope of design-based research. For example, multiple iterations go beyond the resources or the time available to researchers and practitioners.

**Table 9. The significant associations among research topics, research design, and research methods (2003-2012)**

<table>
<thead>
<tr>
<th>Research Topics</th>
<th>Research Design</th>
<th>Research Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Technology or system design supported CSCL</td>
<td>1. Qualitative (AR = 3.7)</td>
<td>1. Experimental (AR = 3.4)</td>
</tr>
<tr>
<td></td>
<td>2. Quantitative (AR = -2.8)</td>
<td>2. Descriptive (AR = 2.0)</td>
</tr>
<tr>
<td></td>
<td>3. Mixed methods approaches (AR = -1.4)</td>
<td>3. Design-based research (AR = 6.9)</td>
</tr>
<tr>
<td>2. Interaction pattern and analysis in CSCL</td>
<td>1. Qualitative (AR = 5.4)</td>
<td>1. Experimental (AR = 5.9)</td>
</tr>
<tr>
<td></td>
<td>2. Quantitative (AR = -3.1)</td>
<td>2. Descriptive (AR = 6.7)</td>
</tr>
<tr>
<td></td>
<td>3. Mixed methods approaches (AR = -2.9)</td>
<td>3. Design-based research (AR = -2.6)</td>
</tr>
<tr>
<td>3. CSCL Evaluation</td>
<td>1. Qualitative (AR = -0.8)</td>
<td>1. Experimental (AR = -0.2)</td>
</tr>
<tr>
<td></td>
<td>2. Quantitative (AR = -0.8)</td>
<td>2. Descriptive (AR = -0.3)</td>
</tr>
<tr>
<td></td>
<td>3. Mixed methods approaches (AR = 1.7)</td>
<td>3. Design-based research (AR = -0.4)</td>
</tr>
<tr>
<td>4. CSCL Practice or applications in natural teaching and community</td>
<td>1. Qualitative (AR = -7.4)</td>
<td>1. Experimental (AR = 8.4)</td>
</tr>
<tr>
<td></td>
<td>2. Quantitative (AR = 5.5)</td>
<td>2. Descriptive (AR = -7.6)</td>
</tr>
<tr>
<td></td>
<td>3. Mixed methods approaches (AR = 2.9)</td>
<td>3. Design-based research (AR = -3.5)</td>
</tr>
</tbody>
</table>

**Discussions**

The trends of CSCL in research topics, research methodology, and technology adoption

This study extended the existed analysis framework and explored the trend of the research topics, research methodology, and technology adoption in CSCL field between 2003 and 2012 based on nine SSCI indexed journals. The results from co-word analysis show that current CSCL research topics can be mainly divided into four categories: technology in support of collaborative learning, interaction pattern and interaction analysis, CSCL assessment and CSCL practices or applications in classroom and community. The thematic evolutions analysis suggests that the research on “CSCL environment,” “assessment of CSCL,” and “interaction pattern and interaction analysis” gained less attention between the two five-year-periods. The researches on “technical affordance and application of social media” and “facilitation of CSCL” have significantly increased between the two consecutive periods. This result corroborated with the expectation of Clark & Mayer (2011) who think that social media such as wiki, blog, Facebook, Twitter should be integrated into the collaborative learning research. Furthermore, previous research revealed that learners typically did not engage in collaboration processes without guidance (Weinberger, Stegmann, Fischer, & Mandl, 2007). Therefore, researchers have paid more attention to facilitating collaborative learning by providing explicit scaffolding (Rienties et al., 2012) and internal and external scripts (Fischer et al., 2013). In the future, concerns related to CSCL support by scaffold, just-in-time feedback, and intervention are very predominant in this field.

Meanwhile, this study examined the trend of research methodology in detail. With respect to research design, there was a significant increase in “Quantitative approach.” More and more researchers adopted a quantitative approach to conduct CSCL research. In addition, Stahl, Koschmann, & Suthers (2006) pointed that experimental methods, descriptive methods and iterative design methods were the three traditional methods in CSCL field. The current study also revealed that “Descriptive research method” was utilized the most and “Design-based research method” was very small in CSCL publications. Also, the proportion of “Design-based research method” has decreased from 4.3% to 3.1% between the two periods. The main reason is that some challenges bound the scope of design-based research. For example, multiple iterations go beyond the resources or the time available to researchers and practitioners.
Another fundamental challenge in conducting design-based research results from assumed dual role of the researchers as designer and researcher (Barab & Squire, 2004). However, the design-based research method can promote the evolution of tools and capture social interactions (Barab & Squire, 2004). As Stahl (2011) suggested, design-based research is specifically appropriate for CSCL research because the software-design process with research is integrated into collaborative learning, which deepens our understanding of how collaborative learning can be supported efficiently. So it is expected that design-based research method should be more adopted in future CSCL studies. Moreover, researchers collected process data, outcome data and miscellaneous data in CSCL investigations. But a significant increase was only found in “Miscellaneous data” between the two periods. This indicated that more and more researchers triangulated between various kinds of data in order to gather substantial evidence.

Furthermore, the current study revealed that “Naturalistic teaching settings” was most often adopted in the last decade. “Higher education” sample group was the most frequently and “Kindergarten” sample groups were the least used during the last ten years. This was in line with the research conducted by Hsu et al. (2012) who found that the higher education group was also used most in the field of technology-based learning. This result implies that undergraduates and graduate students were important sample sources and researchers are access to higher education to obtain data. Future study should also pay more attention to working adults and conduct CSCL research in informal settings. With regard to research learning domain, the results found that the percentages of publications in the “Social science” and “Natural science” learning domain have significantly risen. It is suggested that future study should also conduct research in “Medical Science.”

In addition, “Online collaborative learning” was used most in CSCL research. In terms of technology for supporting collaborative learning, “Social Media,” “Games,” and “Virtual Reality or Augment Reality” have a significant growing trend in the past five years. Technical affordance via social media, games, mobile technologies, and virtual reality or augment reality has continued to dominate the field. This result aligned with the Horizon Reports which identified social media and augment reality technology were likely to enter mainstream use (http://wp.nmc.org/horizon2011/). It was also consistent with the previous report that indicated game-based learning was the core research topics (Kinshuk et al., 2013). Because “Computer Supported Collaborative Learning Environment” became more and more mature, the percentage of publications that described specialized CSCL environment has significantly decreased in the last five years.

The cross analysis between the research design, research methods and research topics

The present study revealed that the research topics were significantly correlated with the research design and research methods. Research on technology in support of collaborative learning utilized most the qualitative approach research and design-based research method. The topic about interaction pattern and analysis used most the qualitative approach research and descriptive method. Mixed method approach was used most in CSCL evaluation research. Research on CSCL practice or applications in natural teaching settings and community adopted the qualitative approach and experimental methods. Therefore, it is suggested that researchers should consider the research topics when they selected research design and research methods.

Conclusions

This study examines the trend of research topics, research methodology, and technology adoption in CSCL field from 2003 to 2012 based on the publications in nine major SSCI journals. The results indicated that CSCL research was dominated by studies that focus on technology in support of collaborative learning, interaction pattern and analysis, CSCL evaluation, and CSCL practice or applications in naturalistic teaching and community. In particular, CSCL research trend has evolved from design CSCL environment, CSCL evaluation, interaction pattern and interaction analysis to technical affordance in CSCL, application of social media, and facilitation of CSCL within the decade. In addition, quantitative research approach has significantly increased in the past five years. More and more researcher collected miscellaneous data as data sources to triangulate the validity of the results. Moreover, the research topics were significantly associated with research design and research method.
The main contribution of this study is twofold. First, this paper extends the previous studies’ frameworks to conduct a thorough and longitudinal analysis of CSCL research topics, research methodology, and technology adoption over the past ten years. The overall review of CSCL researches will be very useful in providing insights into the future research trend. Second, the present study further examines the research trend by comparing the change between the first five years and the second five years. Moreover, the cross analysis between research design, research methods and research topics is conducted to explore the relationships and trends. This kind of approach was rarely reported in previous CSCL review papers. This will also help researchers to identify the research interests and provide educators with good references to make plans in the future. To sum up, the present study contributes a more comprehensive report of CSCL research in contrast to the existing studies.

However, this present study has several limitations. First, this study only analyzed papers of nine leading SSCI journals in the field of CSCL. CSCL conference proceedings and other related journals were excluded. Therefore, the interpretation of the results should be made with caution. Second, the present study only provided the review of research topics, research methodology, and technology adoption in CSCL research. It will be insightful to conduct the further analysis to examine the relationships between instructional interventions and learning outcomes in CSCL field. Finally, this study is limited in analyzing the related research from 2003 to 2012. Further study should be conducted with more current research data.

Acknowledgements

This study was supported by the 2014 Beijing youth project of the Twelfth Five-Year Plan for Educational Science “The empirical study on improving the self-regulated learning abilities in smart learning environment” (CJA14185).

References


An Augmented Reality-based Mobile Learning System to Improve Students’ Learning Achievements and Motivations in Natural Science Inquiry Activities

Tosti H. C. Chiang¹, Stephen J. H. Yang¹* and Gwo-Jen Hwang²

¹Department of Computer Science and Information Engineering, National Central University No. 300, Jung-Da Road, Chung-Li, Tao-Yuan 320, Taiwan // ²Graduate Institute of Digital Learning and Education National Taiwan University of Science and Technology, 43, Sec.4, Keelung Rd., Taipei, 106, Taiwan // tosti.chiang@gmail.com // jhyang@csie.ncu.edu.tw // gjhwang.academic@gmail.com

*Corresponding author

(Submitted July 30, 2013; Revised October 20, 2013; Accepted January 24, 2014)

ABSTRACT

In this study, an augmented reality-based mobile learning system is proposed for conducting inquiry-based learning activities. An experiment has been conducted to examine the effectiveness of the proposed approach in terms of learning achievements and motivations. The subjects were 57 fourth graders from two classes taught by the same teacher in an elementary school in northern Taiwan. The experimental results showed that the proposed approach is able to improve the students’ learning achievements. Moreover, it was found that the students who learned with the augmented reality-based mobile learning approach showed significantly higher motivations in the attention, confidence, and relevance dimensions than those who learned with the conventional inquiry-based mobile learning approach.

Keywords

Teaching/Learning strategies, Elementary education, Cooperative/Collaborative learning, Interactive learning environments

Introduction

Recently, the advancement and popularity of handheld devices and sensing technologies has enabled researchers to implement more effective learning methods (Ogata, Li, Hou, Uosaki, El-Bishouty, & Yano, 2011). Several studies have reported the importance of conducting contextual learning and experiential learning in real-world environments, encouraging the use of mobile and sensing technologies in outdoor learning activities (Chu, Hwang, Tsai, & Tseng, 2010; Hung, Hwang, Lin, Wu, & Su, 2013; Yang, 2006). For example, Chu, Hwang, Huang, and Wu (2008) developed a learning system that guided students to learn about the characteristics and life cycle of plants on a school campus using mobile communication and RFID (Radio Frequency Identification).

Most of mobile learning studies emphasize the adoption of digital learning aids in real-life scenarios (Sharples, Milrad, Arnedillo-Sánchez, & Vavoula, 2009; Ogata & Yano, 2004; Wong & Looi, 2011). However, regarding supplementary mobile learning aids, the interaction between digital learning aids and the actual environment needs to be emphasized to enable students to effectively manage and incorporate personal knowledge (Wu, Lee, Chang, & Liang, 2013). For example, it is explicated that students can select a virtual learning object from the actual environment using a mobile learning aid, which allows them to obtain a first-hand understanding of the learning environment and, subsequently, increases their learning motivations and experiences. Such a learning support technology is achievable through the use of Augmented Reality (AR), which combines human senses (e.g., sight, sound, and touch) with virtual objects to facilitate real-world environment interactions for users to achieve an authentic perception of the environment (Azuma, 1997). For example, users who employ mobile devices with AR facilities to seek a target building on a street are able to see additional information surrounding individual buildings when they browse the buildings via the camera of their mobile device. Researchers have documented the potential of employing such facilities to assist students in learning in real-world environments in comparisons with traditional instructions (Andujar, Mejias, & Marquez, 2011; Chen, Chi, Hung, & Kang, 2011; Kamarainen, Metcalf, Grotzer, Browne, Mazzuca, Tutwiler, & Dede, 2013; Platonov, Heibel, Meier, & Grollmann, 2006), which showed that AR technology contributed to improve academic achievement compared to traditional teaching methods.

On the other hand, numerous educators have contended that computer technology cannot support the learning process entirely; instead, the primary function of computer technology involves a knowledge building tool for students (Jonassen, Carr, & Yueh, 1998). Effective learning strategies remain the most crucial factor for increasing
learning motivation. Therefore, effective learning strategies supplemented with appropriate computer technology can greatly enhance learning motivation (Chu, Hwang, & Tsai, 2010; Jonassen, 1999; Hwang, Tsai, Chu, Kinshuk, & Chen, 2012; Jonassen et al., 1998). Previous studies have highlighted that inquiry-based learning strategies supplemented by computer technology in a scenario-based learning environment can effectively increase learning motivation (Shih, Chuang, & Hwang, 2010; Soloway & Wallace, 1997). Inquiry-based learning strategies are student-centric knowledge exploration activities; the teacher serves as a guide, employing structured methods that train and encourage students to learn proactively (Hwang, Wu, Zhuang, & Huang, 2013; Soloway & Wallace, 1997). When students acquire the methods for problem-solving, they use the obtained information to establish a hypothesis or to plan solutions to the problem (Looi, 1998).

Consequently, in this study, an innovative learning approach is proposed to support inquiry-based learning activities with mobile AR. During the activities, the learning system guides the students to complete their learning tasks and provides learning supports by sensing their locations. Through the mobile AR technology, the learning system is able to present differing learning contents on the handheld devices based on individual students’ learning scenarios, such that the students can interact with the learning content to gain relevant knowledge (Yang & Chen, 2008). Finally, the entire learning process is recorded and uploaded, allowing students to share their experiences and perceptions with their peers.

To evaluate the effectiveness of the proposed approach, the following research questions are investigated:

- Do the students who learn with the inquiry-based mobile AR approach have better learning achievements than those who learn with the conventional inquiry-based mobile learning approach?
- Do the students who learn with the inquiry-based mobile AR approach reveal higher learning motivations than those who learn with the conventional inquiry-based mobile learning approach?
- Is there a significant difference between the cognitive loads of the students who learn with the inquiry-based mobile AR approach and the conventional inquiry-based mobile learning approach?

### Literature review

#### Inquiry-based Learning

Inquiry-based learning is a learning activity that involves the teacher encouraging students to proactively hypothesize, explore, validate, categorize, explain, and discuss everyday situations or problems encountered. Through hypothesizing, exploring, and observing, students develop advanced social interaction skills and a higher level of thinking. Inquiry-based learning enables students to not only develop a deeper level of thinking when encountering learning situations but also to learn how to implement the process of learning (Price, 2001). Lim (2004) contended that during the process of conducting learning tasks, online inquiry-based learning allows students to develop confidence to participate in activities, cultivate teamwork abilities, and feel greater responsibility for controlling their learning progress. Creedy, Horsfall and Hand (1992) asserted that traditional learning methods are merely knowledge transfer activities, that is, instructors transfer their knowledge to students, whereas inquiry-based learning allows students to learn proactively rather than passively receiving knowledge. Colburn (2000) defined inquiry-based learning as a method that comprises many open, student-centered, and hands-on approaches, including structured inquiry, guided inquiry, open inquiry, and the learning cycle. These approaches enable students to identify connections on various levels and to integrate and collate information. The instructor then provides the students with relevant concepts, and after this knowledge is assimilated, students are able to apply it in other contexts.

The traditional approach informal education has been criticized for creating artificial classroom contexts where the learning activities and resources become divorced from their meaning in real life situations (Herrington, Reeves, & Oliver, 2010). The advocates of authentic learning argue for the creation of more meaningful learning situations. Authentic learning requires that the contexts used for learning reflect real world contexts where the skills and competencies will be deployed. Mayer (2001) also emphasizes the learners' learning outcomes in integrated text and the illustrator will better than divorced. These AR-based inquiry scenarios can cover a wide range of domains from observation and interaction through to accessing real scientific instruments over the Web to conduct more realistic experiments.

Edelson, Gordin and Pea (1999) attested that the inquiry experience is beneficial for students to correct their
scientifical knowledge. In recent years, inquiry-based learning has been broadly applied in nursing education (Akinsanya & Williams, 2004; Finn, Fensom, & Chesse-Smyth, 2010). Inquiry-based learning is also applied in the social sciences (Ahmed & Parsons, 2013; Bowman, 2012). Lakkala, Lallimo, and Hakkarainen (2005) combined history classes with inquiry-based theory for implementation in 12 elementary and junior high schools. Shih et al. (2010) integrated inquiry-based theory with mobile devices to assist students in understanding the culture associated with temples. Chang, Wu and Hsu (2013) also combines mobile AR technology and pedagogical inquiry activities in a socioscientific issue's context is effective in terms of promoting students' understanding of the science content for ninth-grade students.

Augmented reality

AR is a technology that allows users to combine real-life sensory experience with digital environment perceptions (Azuma, 1997). According to Azuma, Baillot, Behringer, Feiner, Julier and Maclntyre (2001), the three characteristics of AR are (a) real and virtual objects incorporated into reality; (b) collaboration between real and virtual objects, and (c) real-time interaction between real and virtual objects.

Display and positioning technologies fulfill the three criteria of AR; differing sensory input is received through environmental changes and, subsequently, the interaction and changes between senses and scenarios are adjusted. Sherman and Craig (2003) further explained that the presentation of sensory characteristics is only achievable through the level of environmental reality and activity immersion; therefore, differing levels of sensory experience can be achieved through different display and positioning technologies. Furthermore, if the augmented information of the real object is comprehensive, the sensory experience and knowledge transmission will be accurate; conversely, if the augmented information is incomplete, the sensory input and knowledge transfer will be inaccurate (Wu et al., 2013). Considering these points, the greatest challenges educators face regarding using AR is how to increase learning motivation by enhancing the students' sensory experience. The second characteristic of AR is the cooperation between real and virtual objects: The aim of mobile AR is to integrate digital data into the real environment to provide users with an immersive sensory experience (Hwang, Yang, Tsai, & Yang, 2009; Hwang, Tsai, & Yang, 2008; Yang, Okamoto, & Tseng, 2008). Therefore, the virtual object must be presented accurately in the real geographical location; this is the essence of the navigating mechanism. In a virtual or augmented scenario, the navigating mechanism does not actively interact with the user or provide accurate information; instead, it passively interacts with and provides information to the user (Kye & Kim, 2008; Burigat & Chittaro, 2007). The final characteristic is that real and virtual objects must interact in real-time: the mobile AR scenario supports three types of interaction. The first is the interaction between the student and the learning content. Numerous previous studies have indicated that this type of interaction increases students’ cognitive and learning abilities, such as comprehension, memory, and imagination (Dalgarno, 2004). The other two types are the interaction between the student and the learning aids, and the interaction between students. These two types of interaction enable students to identify solutions to problems in the scenario through cooperation and teamwork.

Educators and researchers anticipate applying emerging technologies, such as AR and multiple-user VR, to teaching and learning activities (Bower, 2008; Dalgarno & Lee, 2010; Dunleavy et al., 2009; Kye & Kim, 2008). The sensory experience and interaction and guiding functions of these technologies can improve students’ learning satisfaction and enable them to structure their knowledge and complete the learning tasks (Dalgarno & Lee, 2010; Dunleavy et al., 2009). Furthermore, a number of researchers have proposed that mobile AR devices have unique applications in education, such as improving the success rate of physical interaction-related learning tasks and supporting memory-related learning activities (Chien, Chen, & Jeng, 2010; Dunleavy et al., 2009).

Augmented reality-based mobile learning approach

The AR-based mobile learning system was developed using JAVA for the website, Oracle for the database, and Xcode for the iPad mini devices. Figure 1 shows the structure of the system, which consists of a location, a camera, image editing, a digital compass, a three-axis gyro, an accelerometer and an AR display module.

The location module is able to detect the GPS location of the students, guide them to find the target ecology areas, and show them the corresponding learning tasks or related learning materials. The camera and image editing modules
are able to capture the image from the authentic environment and to annotate and comment on the image of the observed objects when students investigate different characteristics of learning objects. The edited images, comments, annotations, and GPS location data are uploaded to the media server by WIFI communication networks. The digital compass and the three-axis gyro are able to distinguish the direction and relative positions of the students and learning objects. The accelerometer module allows students to shake their iPad to catch the authentic image through small micro electro-mechanical systems. Finally, the AR display module, according to the different locations of the learning objects, shows the integrated images consisting of the target learning objects, edited images and relevant information, on the screen of the mobile device.

Figure 1. System structure of the interactive AR-based mobile learning

Figure 2 shows the basic functions of the interactive AR-based mobile learning system, which consists of an AR-based mobile learning function, an online chat room function and an investigated portfolio function. The AR-based mobile learning function allows students to learn about the target learning objects, to link to the supplementary materials, to capture their observations, to annotate and to comment on the images, and to browse other students’ observations. The online chat room function allows students to discuss their investigation immediately from different locations. The investigated portfolio function collects each user’s observed portfolio, participated portfolio and reflected portfolio.
Figure 2. Interface of the interactive AR-based mobile learning

Figure 3. Example of providing AR information
Figure 3 shows an illustrative example of applying the interactive AR tool. The AR tool in the mobile device integrates images and a description on different layers according to different distances and directions. In the upper right corner of the screen, there is a “rotate” function in order to make it easier for students to alter the overlapping layers. Image-based discussion can prompt students to share more observations and reflections by pressing the image to leave comments. The distance on the image can guide students in the right direction to walk toward, observe, and touch the learning object.

Figure 4. Students’ activity for AR-based mobile learning

Figure 4 shows the flow of the AR-based inquiry-based learning system, which is implemented based on the 5-step learning strategy proposed by Bruce and Bishop (2002):

(1) Ask: The instructor first defines the learning objects and allows students to search for them to cultivate their proactivity. The purpose of this step is to enable students to search for the learning objects, which are then continuously redefined throughout the cycle of the inquiry-based learning. First, students must define the learning objects to be explored. Thus, the instructor provides background knowledge of the learning objects in the class, and by piquing the students’ interest in these objects, their learning motivation is greatly enhanced. Then the instructor guides the students to the set learning scenario using mobile AR, allowing the mobile AR to lead the students to the target learning objects.

(2) Investigate: In this step, students are naturally guided to continuously investigate the learning content by their curiosity. Students can reference the learning aids to understand various aspects of the learning content. When the content is understood, they can redefine the learning object or simplify it by dividing it into smaller components. When students have reached the learning object, the instructor can guide them in observing the environment using the questions presented by the handheld device. Through observation, the students can examine whether the information on the learning object is identical to their knowledge and then, using mobile AR, extend their learning to other concepts of the object. Students can use these extended concepts to conduct deeper observations and exploration.
During the observation process, students can use the camera function to take photographs of the learning object and share their observations with their peers (see Figure 3). They can also employ the footnote function to record their thoughts and examine the accuracy of the results with their peers. Using this technology, students can not only perform inquiry-based learning more easily and increase their understanding of the learning object, but they can also develop teamwork skills that facilitate problem-solving during the learning process (Shih et al., 2010). In addition, students can instantly upload their thoughts and observations, enabling the instructor to constantly monitor their learning progress.

(3) Create: After the inquiry process, the instructor then uses the learning progress system to discuss the concepts related to each learning object with the student. During this process, the instructor redefines the learning objects, enables the students to share their experiences, and stimulates a discussion of relevant ideas. These in-depth cognitive processes allow the students to assimilate the information of the learning object into their own knowledge. This step shows that the role of the instructor is to guide students in developing their knowledge without participating in the actual process. After the completion of learning object investigation, the students form connections among the learning content to create new internalized knowledge.

(4) Share: Following the completion of knowledge construction, students can use the learning progress system to share their learning experiences and perceptions. Through sharing, students can reflect on whether they should adjust their methods for understanding the teaching material and examine whether their ideas correspond with those of their peers. Furthermore, the progress data captured by the learning progress system not only allows effective management of students’ learning objectives but also enables the instructor to identify whether the students are experiencing difficulties during the learning process. Instructors can monitor the students’ learning condition based on their progress, assess whether they should adjust their teaching method, and understand why particular teaching methods cause learning difficulties for students.

(5) Reflect: After sharing their learning experience with their peers, the students reflect on and reconsider their newly acquired knowledge on a deeper level. Reflection is a critical aspect of the learning process (Chi, de Leeuw, Chiu, & Lavancher, 1994), significantly influencing students’ knowledge comprehension and memory. Through reflection exercises, students can revise, examine, and correct their knowledge and perceptions. During the reflection process, knowledge is more easily understood and better comprehended, which improves learning achievement. Reflection also promotes learning independence because it highlights the various levels of a problem, allowing students to examine, evaluate, and understand their thoughts. This increases their involvement in the learning process, enhances their learning motivation, and enables students to become positive, proactive, and responsible self-reflecting individuals. Students can rethink the initially defined learning objects and the direction of inquiry, as well as confirming the accuracy of their conclusions.

Experiment design

The experimental material used in this study was a fourth grade natural science unit on aquatic animals and plants, which was divided into four sections: water habitats, different types of aquatic plants, different types of aquatic animals, and the secret of aquatic plants. Each section contains learning themes, for example, water habitats is comprised of the two themes of natural habitats and manmade habitats; natural habitats include lakes, rivers, marshes, coastal intertidal zones, and lake intertidal zones; manmade habitats include ponds, dams, and irrigation ponds. Water plants comprise the following four themes: emergent plants, submerged plants, floating-leaf plants, and floating plants. The inquiry-based learning activity was developed based on the five-step design criteria proposed by Bruce and Bishop (2002), that is “ask,” “investigate,” “create,” “share” and “reflect.”

Participants

The participants of this experiment were fourth grade students from an elementary school in Northern Taiwan. A total of 57 students were included in this study, ranging between 9 and 10 years of age. These students were from two classes; one class was set as the experimental group, and the other was the control group. The same instructor was responsible for both classes.
Experimental procedure

Figure 5 shows the experimental procedure. During the learning activity, the instructor used 90 minutes to introduce aquatic plants. Following that, the students took a 30-minute pre-test, which aimed to evaluate whether the two groups of students had an equivalent basic prior knowledge of the natural science course content. The students in both groups were trained to operate the mobile learning devices before the 120-minute inquiry-based investigation. During the learning activity, the students in the experimental group learned with the AR-based mobile learning approach. On the other hand, those in the control group learned with the inquiry-based mobile learning approach; that is, the instructor provided mobile devices and engaged the students in inquiry-based learning activities for investigating the issues of water hyacinths and their distribution.

The learning materials included text descriptions and pictures of the plant characteristics. For example, water hyacinths are aquatic plants and are pale purple in color. After presenting a film on water hyacinths, the instructor asked the students to examine the plant characteristics based on the information in the film and to record their thoughts on their mobile devices. After exploring the plant characteristics, the instructor let the students discuss concepts related to water hyacinths and encouraged them to discuss their ideas and determine whether their knowledge corresponded with that of their peers using their mobile devices. After the basic concepts regarding water hyacinths were understood, the instructor let the students present other plants that are similar to water hyacinths, such as water lilies and duckweed. Thus, the students could identify the differences between these plants and water hyacinths, enabling them to learn the characteristics of other plants and reinforcing their knowledge of water hyacinths.

![Figure 5. The experimental procedure](image)

Measuring tools

To assess the students' learning achievements, a pre-test was conducted to ensure that the two groups of students had equivalent prior knowledge before the learning activity. It consisted of thirty multiple-choice items with a perfect score of 100. Moreover, a post-test was conducted for assessing the students' learning achievements after the learning activity. It consisted of thirty multiple-choice items for assessing the students' knowledge for identifying and differentiating the plants on the school campus with a perfect score of 100. Both the pre-test and the post-test were developed by two experienced teachers who had more than 5 years experience in teaching the natural science course.
The learning motivation questionnaire was modified from the measure developed by Keller (2010) based on the ARCS (Attention, Relevance, Confidence, and Satisfaction) model of motivational design (Keller, 1987). A total of 36 questions in four dimensions were employed and assessed using a five-point Likert scale; that is, the design of the learning activities must attract the students’ attention (Attention), the learning activities and materials have to be relevant to the students (Relevance), the students have to be confident with the learning activities (Confidence), and the students need to feel satisfied after completing the learning activities (Satisfaction). The Cronbach’s α values of the four dimensions were 0.80, 0.78, 0.65 and 0.82, respectively.

The cognitive load survey developed by Sweller, van Merriënboer and Paas (1998) was used to measure the cognitive load of individual students. The main purpose of the questionnaire was to assess whether using a tablet computer as a learning tool generates cognitive load. The questionnaire covers four questions; two measure the students’ mental load and the other two measure their mental effort using a five-point scale. The Cronbach’s α values of the two dimensions were 0.845 and 0.85, respectively.

Analysis and results

Learning achievement

Before the experiment, the two groups took a pre-test to ensure that they had equal abilities in this subject before the learning activity. The means and standard deviations of the pre-test were 46.46 and 13.220 for the experimental group, and 44.97 and 15.546 for the control group. The t-test result showed that these two groups did not differ significantly (t = 0.391, p > .05); that is, the two groups of students had statistically equivalent abilities before learning the subject unit.

After participating in the learning activity, the two groups of students took a post-test. The t-test result shows that the average learning achievement of the experimental group was significantly better than that of the control group (t = 2.046, p < .05), as shown in Table 1.

From the above results, it is concluded that the mobile AR approach is helpful to the students in improving their inquiry-based learning achievements.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Std. error</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test</td>
<td>Experimental group</td>
<td>28</td>
<td>80.14</td>
<td>8.763</td>
<td>1.656</td>
<td>2.046</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>29</td>
<td>73.93</td>
<td>13.703</td>
<td>2.545</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.

Learning motivation

Table 2 shows the t-test result of the learning motivation of the two groups. The t-test result showed that the difference between the learning motivations of the two groups was significant (t = 2.99, p < .01); moreover, the mean value of the experimental group (i.e., 4.05) was higher than that of the control group (i.e., 3.63), implying that the mobile AR approach can promote students’ motivations in inquiry-based learning activities.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Experimental group</td>
<td>28</td>
<td>4.05</td>
<td>0.07</td>
<td>2.99**</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>29</td>
<td>3.63</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

**p < .01.

Table 3 further shows descriptive statistics for the four subscales of the ARCS model used to describe motivation. It was found that the mean values for attention, confidence and relevance of the experimental group were significantly higher than those of the control group.

---

Table 1. t-test results of learning achievement of the two groups

Table 2. t-test results of the learning motivation of the two groups
The primary objective of the ARCS model of motivation was to assess the students’ attention regarding the learning material and learning activities. In the interviews, the students in the experimental group indicated that the virtual learning material presented through the mobile AR scenario enabled them to focus on exploring the characteristics of plants because they did not need to spend time linking the digital supplementary materials or instructions from the mobile learning system to the real-world contexts. The students also indicated that mobile AR allowed them to access learning materials and learning tasks immediately, regardless of their real-world locations, and hence they would proactively attempt to understand the learning content provided by the learning system and the surrounding real-world learning objects.

The relevance hypothesis of the ARCS model of motivation was used to assess the relevance of the learning tasks and the provided materials. It was found that the students in the experimental group gave significantly higher ratings than those in the control group in this dimension since the mobile AR approach provided immediate and relevant information and guidance to the students based on their real-world locations and contexts, which was highly beneficial to them for their outdoor observations and learning. On the other hand, those in the control group needed to search for relevant information regarding their learning contexts on their own, which lowered their learning motivation to some extent.

The confidence hypothesis of the ARCS model of motivation was used to assess the students’ confidence regarding the learning activities and their level of anticipation. In the interviews, the experimental group students indicated that mobile AR could actively provide support or guidance related to the real-world contexts. Such an integration of real-world contexts and digital-assistance was helpful to them in facing the challenges of the learning tasks, and hence could promote their confidence regarding the learning activities.

The satisfaction hypothesis of the ARCS model of motivation was used to assess the students’ feeling of satisfaction after completing the learning activities. As both groups of students were equipped with the same mobile devices with digital supplementary materials regarding the aquatic plants, all of the students showed a high level of satisfaction in using the mobile devices to learn, resulting in no significant difference in this dimension.

**Cognitive load**

The aim of using the cognitive load measure was to evaluate whether the students’ performances were affected owing to improper educational settings, including the difficulty levels of the selected learning materials and the learning strategies adopted. The experimental results showed that the means and SD values were 2.02 and 0.83 for the experimental group, and 2.27 and 0.79 for the control group. Moreover, the t-test results for the cognitive loads of the two groups were $t = -1.168$ and $p > .05$, indicating that the two groups’ cognitive loads did not differ significantly.

The cognitive load measure consisted of two dimensions: mental load and mental effort. An in-depth analysis was further conducted on these two dimensions. Mental load refers to the internal aspects of cognitive load; specifically, when students face a large amount of learning content or difficult content beyond their information processing abilities or knowledge levels, they may perceive an excessive cognitive load. For mental load, the experimental group’s mean = 2.23 and $SD = 0.98$, while the control group’s mean = 2.24 and $SD = 0.90$. The t-test results showed that $t = -0.037$, $p > .05$, which indicates no significant difference between the two groups’ mental load. This finding is reasonable since both groups of students were situated to complete the same learning tasks with identical
supplementary materials.

Mental effort refers to whether the students must exert more mental effort to understand the learning materials. Table 4 shows the experimental group’s mean = 1.80 and SD = 0.91, while the control group’s mean = 2.29 and SD = 0.99. The t-test analysis results showed that t = -1.949, p > .05, indicating that no significant difference was found. From the means of the two groups, it was found that the mental effort of the students who learned with the mobile AR approach showed slightly lower mental effort, which might be attributed to the use of the AR technology in helping them link the real-world contexts with supplementary materials at the right place and the right time.

Table 4. t-test results of the two subscales of cognitive load of the two groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental load</td>
<td>Experimental</td>
<td>28</td>
<td>2.23</td>
<td>0.98</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>29</td>
<td>2.24</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Mental effort</td>
<td>Experimental</td>
<td>28</td>
<td>1.80</td>
<td>0.91</td>
<td>-1.949</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>29</td>
<td>2.29</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

**Discussion and conclusions**

In this study, a mobile AR approach is proposed for conducting inquiry-based learning activities. A learning system was developed based on the proposed approach and an experiment has been conducted to evaluate the effectiveness of the approach in an elementary school natural science course. The experimental results show that the mobile AR approach is able to improve students’ learning performance in inquiry-based learning activities owing to the use of AR technology in linking the real-world contexts with the digital learning resources at the right place and the right time. Such a result can be explained by spatial and temporal continuity principles of the multimedia design theory proposed by Mayer (2001) and Mayer and Moreno (2003); that is, learning from learning scenarios that present relevant materials (e.g., images, texts, videos) in a well integrated and organized form can avoid creating incidental cognitive load, and hence benefits students in improving their learning performance. Similar findings have been reported by Liu, Lin, Tsai and Paas (2012) in web-based learning activities. When learning with the AR-based mobile learning system, the students learned from the scenarios that presented the real-world targets and the supplementary digital materials in an integrated and organized way. On the other hand, in a traditional instruction approach or a conventional mobile learning approach, the real-world targets and the corresponding materials are presented separately and asynchronously. When observing the real-world targets, the students need to read the corresponding materials from a mobile device or a printed sheet and put lots of efforts on organizing the information by themselves, which prevents them from viewing the learning targets and thinking in a higher order manner.

In addition, the experimental results indicate that the experimental group students gained significant learning motivation for attention, relevance, confidence and a high level of satisfaction in using the AR-based mobile devices to learn. According to the interviews, students in the experimental group thought the AR-based mobile learning felt interesting and useful for assisting them to learn. Moreover, students pointed out that using the AR-based mobile learning to conduct inquiry-based learning can not only provide opportunities to practice, but also to engage in enjoyable experiences for assisting Inquiring. Students were excited and gained a feeling for the interesting in real-world environments when taking photos, making fancy notations on photos, sending to the AR-based learning system from different places or sharing their comments with classmates. Therefore, the students had a positive learning motivation toward using AR-based mobile learning to aid natural science learning and were also satisfied with its effectiveness.

Although the AR-based mobile learning system benefited the students in this application, there are some limitations to be noted. First, the GPS accuracy of the mobile devices limits the display of the location of the learning objects; therefore, when designing a learning task, the teachers need to consider the size of the learning objects and the distance between them. Moreover, to provide instant hints or learning guidance to individual students, the teachers need to spend time developing learning processes for evaluation purposes, and digital learning materials to provide learning supports.

In the near future, we will try to apply this approach to other mobile learning applications, including the natural science courses and local culture courses of elementary and high schools. Moreover, we plan to explore the
behavioral patterns of an online knowledge-sharing discussion activity for inquiry-based learning courses and explore the performance of using AR-based mobile learning for different learning styles or different cognitive styles. Adopting different learning devices, such as Google Glass, can also become one of our research directions.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract numbers NSC 102-2511-S-008-007-MY2 and NSC 99-2511-S-008-006-MY3.

References


Phillips (Eds), *Beyond the Comfort Zone: Annual conference of the Australasian society for computers in learning in tertiary education (ASCILITE)* (pp. 240-248). Perth, Western Australia: ASCILITE.


2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, JhongLi, Taiwan.


A Multi-modal Digital Game-based Learning Environment for Hospitalized Children with Chronic Illnesses

Jui-Chih Chin¹ and Mengping Tsuei²*

¹University of Taipei, Taiwan // ²National Taipei University of Education, Taipei, Taiwan // jcchin@utaipe.edu.tw // mptsuei@tea.ntue.edu.tw

*Corresponding author

(Submitted August 16, 2013; Revised November 19, 2013; Accepted January 27, 2014)

ABSTRACT

The aim of this study was to explore the digital game-based learning for children with chronic illnesses in the hospital settings. The design-based research and qualitative methods were applied. Three eight-year-old children with leukemia participated in this study. In the first phase, the multi-user game-based learning system was developed and implemented during the first iteration. Children could create their own or co-construct narratives and play mathematics games in the multi-user system. Then, we developed the multi-modal digital game-based learning activities in the second iteration. Children showed highly motivation to engage in learning activities. A list of key design features related to the digital game-based learning for children with chronic illnesses emerged from the data. The results supported that the multi-modal digital game-based learning provided the social interactive processes and learning motivation, which effectively served the learning and psychosocial needs of chronically ill children.

Keywords

Children with chronic illnesses, Digital game-based learning, Bedside teaching, Elementary education

Introduction

Digital games are rapidly becoming important tools for education, training, and even healthcare. Many people use these media for entertainment purposes and to escape the difficulties of social life (Sherry, 2004). Pivec (2007) indicated that the Digital Game-Based Learning (DGBL) model provides a complex learning opportunity, increases learners’ motivation, and offers a different mode of interaction and communication. Recently developed health-related games have been shown to effectively facilitate healthy behaviours, such as healthy lifestyle habits, behavioural modifications, self-management of illnesses and chronic conditions, and physical activity (Ferguson, 2012). The effectiveness of DGBL is not limited to the general student population, but extends to students with special needs. DGBL has been shown to improve the motivation of students with learning disabilities (Ke & Abras, 2012), the attention of children with cognitive disabilities (Rezaiyan, Mohammadi, & Fallah, 2007), the psychomotor skills of slightly mentally disabled children (Karal, Kokoç, & Ayıyıldız, 2010), and the social problem-solving skills of students with attention deficit hyperactivity disorder (Goldsworthy, Barab, & Goldsworthy, 2000).

Educators face the challenge of meeting the individual needs of students with very diverse backgrounds and educational requirements. Given the high incidence of childhood illness, supportive programmes to maintain students’ motivations to learn during extended absences due to chronic conditions are needed (Leger & Campbell, 2008). However, young children with chronic illnesses have attracted little attention and support from educational communities (Wilson-Hyde, 2009). Students with chronic illnesses share the need for equal access to the same educational outcomes, academically and socially, as their healthy peers (Shiu, 2001). DGBL is a valuable instructional strategy for increasing students’ motivation and providing opportunities for interaction and communication among students (Pivec, 2007). Despite increased interest among educational professionals in the use of digital games to support children’s learning, research on examining the potential of DGBL for children with chronic illnesses is extremely limited.

This study explores the potential of DGBL to facilitate learning motivation and social interaction for hospitalized children with chronic illnesses. The study question is: what are the features of DGBL that promote learning motivation and prompt social interaction for children with chronic illnesses? By drawing upon observations and discussion with ill children, their parent, and a bedside teacher, the current study illustrates the needed support for the hospitalized children during the DGBL activities.
Background

An estimated 2% of children in North America have severe chronic conditions, and an additional 3% have multiple chronic conditions (Newacheck & Stoddard, 1994). The incidence of illnesses such as asthma (Peat et al, 1994), diabetes (Kelly, Russell, Jones, & Byrne, 1994), and leukaemia (American Cancer Society, 1994) has increased in children. In Taiwan, there were about 0.06% of K-6 graders with severe chronic conditions (MOE, 1995).

Consideration of the special needs of students with chronic illnesses must take into account academic and social aspects of the school setting (Shiu, 2001). Chronic medical conditions severely limit children’s ability to participate in normal educational and recreational activities. These students have no learning impairment, but their ability to regularly attend school is limited (Ruland, Starren, & Vatne, 2007). The Ministry of Education in Taiwan required all children with chronic illness to accept beside teaching or home schooling services while they were absent from school. However, a family can only receive one weekly session (Tsai, 2006). The hospitalized children often struggle to understand and complete school work without assistance. The lack of support impedes learning opportunities and hardly sustains their learning motivation. An empirical study on children with cancer indicated that bedside instruction received in hospital were bored and did not resemble what children would have done in school (Charlton, Pearson, & Morris-Jones, 1986).

Children with chronic illnesses also face many challenges, such as repeated invasive procedures and treatments, changes in daily routine and relationships with family members and peers, and increased emotional insecurity and disturbance. A previous study indicated that children with cancer may be more knowledgeable and mature than healthy children in some respects due to their experience with the illness, but that they are more immature in other respects due to dealing with a life-threatening illness, treatment disrupts some aspects of their daily lives, and dependence on others (Ruland, Slaughter, Starren, & Vatne, 2006). Children with long-term or chronic illnesses often have difficulty maintaining social contacts with their peers, which increases their feelings of social isolation and loneliness (Lightfoot, Wright, & Sloper, 1999). The lack of social support makes these children more vulnerable than children in general (Shiu, 2004). Issues of how technologies may be utilized in a hospital context to provide real-time, interactive, and educational experiences for young students with chronic illnesses deserve more concerns.

Technology for students with chronic illnesses

Technology is an efficient means for children who are physically constrained to maintain social connections. Isolated children may communicate with others or engage in activities via web-based technology, which presents various opportunities to form peer groups or virtual communities. Evidence indicates that computer technology can address the psychological needs of children with chronic illnesses.

Bers (2001) created the Zora system, a three-dimensional (3D) multi-user computer environment, to help children who had undergone organ transplantation form virtual communities to share their experiences and combat the isolation created by their medical situations. This system allows paediatric patients to build virtual rooms, converse with others in real time through an avatar, and post messages or stories. By engaging in these activities in virtual communities, patients develop senses of belonging and group identity. Participating patients found that the Zora system helped them make friends and feel connected with the social world (Bers, 2001). They also responded positively to synchronous as well as asynchronous communication. Bers (2001) stated that internet technologies that enable the formation of virtual peer communities present unique opportunities for possibly isolated children.

Similarly, research also explored the feasibility of technology-mediated interaction for sustaining academic continuity between children with chronic illness and their teachers at school. Fiore et al (2008) developed an interactive community platform that integrates various synchronous communication components in a virtual 3D environment, allowing children with chronic illnesses to communicate with their classrooms. These components include two-way, synchronous video streams of lessons; digital diaries and timetables; asynchronous lessons; homework exchange; exercises, tests, and marks; and a virtual playground. Implementation of this platform dramatically improved social bonds between ill children and their peers and teachers (Fiore et al, 2008).

Fels and Weiss (2001) developed the PEBBLES (Providing Education by Bringing Learning Environments to Students) video-mediated communication application to link hospitalised children with their regular classrooms. One component of the system was located in the classroom, and the other component was located in the dialysis unit of a
hospital. Analysis of video data obtained during a 6-week case study revealed that the participating hospitalised child spent most in-class time on academic tasks. Use of the PEEBLES system increased her classmates’ understanding and awareness of what she endured, as well as positive interactions with her (Fels & Weiss, 2001).

The aforementioned studies demonstrated technological possibilities for linking the chronically ill children to the learning and interactive opportunities with school teachers and peers. However, teachers-participants have highlighted that they struggle to find additional time beyond their regular workloads in a sustainable way (Fels & Weiss, 2001; Fiore et al, 2008). The issues about such technology use, however, extend beyond the mere technical potential of new technologies to their pedagogical implications and feasible implementation.

Despite the potential effectiveness of technological support for children with chronic illnesses, only a few older children use it frequently (Fiore et al, 2008). However, the overall prevalence of childhood chronic conditions is similar in younger (aged < 10 years) and older (aged 10–17 years) children (Newacheck & Taylor, 1992). Hospitalised children have reported that computers are their favourite educational tools (Lombaert, Veevaete, Hauttekeete, & Valkcke, 2006). Play is essential in children’s lives, and computer games provide a vehicle for play not only as a diversion, but also as an integral part of their education and social lives (Prensky, 2001). Our pilot study indicated that the synchronous communication enhanced healthy children’s sense of agency and facilitated a supportive partnership with other players. Perceiving self as an active learner and having companionship are in tune with the needs of hospitalized children. The effectiveness of DGBL for chronically ill children has not been well documented. Therefore, this paper investigated the potential of DGBL for supporting social and educational engagement among students with chronic illnesses.

Methods

This study seeks to initiate a discussion about the design and implementation of DGBL for hospitalised children with chronic illnesses following design-based research and a grounded theory approach.

Design-based research attempts to combine theory-driven design with empirical analyses of learning environments using a multi-step iterative cycle proceeding from design to evaluation and refinement to connect interventions with outcomes (Collins, Joseh, & Bielaczyc, 2004). The grounded theory approach (Glaser & Strauss, 1967) was used to inform the processes of data collection and analysis during the iterative cycle. This study involved two iterations: in the first iteration, we developed the “Kala Forest” multi-user DGBL system and implemented it with ill children; in the second iteration, we refined the model to create a multi-modal DGBL system for children with chronic illnesses.

Participants

There were three chronically ill children participating in this study who were referred from Taipei City West Special Education Resource Center. The children were hospitalized in a children’s hospital in Taipei. We obtained consent from the parents and from the bedside teacher before data collection.

Two 8-year-old boys with leukaemia (John and David; all names are fictitious) and one with neuroblastoma (Ted). These children indicated that they missed the classroom experience; John had attended a public elementary school for 1 month, and Ted and David had never been to elementary school due to their medical conditions. These children were repeatedly hospitalised. The mean duration of hospitalisation was 19.3 ± 12.5 days, during which the special educational service centre administered by the Taipei city government provided bedside teaching services. When they were not hospitalised, the children received home schooling.

The children’s parents provided them with notebook computers and wireless connections while they were hospitalised. Additional information about the participants is provided in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Participants’ information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
</tr>
<tr>
<td>John</td>
</tr>
<tr>
<td>Ted</td>
</tr>
</tbody>
</table>
at the first graders’ level. Observed behaviors showed that he had introvert personality. Great parental supports for reading experiences. He can read the instruction and play with game independently.

David Mathematics, reading and writing skills reported by the bed-side teacher are at the second graders’ level. The parents insisted living by the same daily schedule as the school when David was in the hospital. David’s mother taught him most of curriculum content. He needs one-to-one instruction for understanding the game rules. David participated in this study one month later than John and Ted did.

Data collection

Data collection involving children with chronic illnesses adds dimensions of complexity and challenge. The data collected in this study included observational field notes and artifacts during the implementing DGBL sessions and interviews with participating children, their parents, and the bedside teacher.

First iteration: Implementation of the Kala Forest DGBL system

In the first iteration of this study, we investigated whether the Kala Forest DGBL system contributed to the social and learning processes of children with chronic illnesses in a hospital setting.

Kala Forest: A multi-user digital game-based learning system for children with chronic illnesses

To accommodate the needs of children with chronic illnesses, we sought to create an easy-to-use system that was not cognitively or emotionally demanding. Although there were some commercial systems available, these systems were not suitable for the ill children. For example, most of the drawing systems did not provide the graphical objects related to hospitalized children’s life experiences. Moreover, children with chronic illnesses were sensitive and alert to strangers. The existing online systems open to all public users could not meet their needs. Therefore, we intended to develop the multi-user DGBL system which could engage the ill children in playing, learning, and sharing. The conceptual design of the Kala Forest system was as follows:

- Enhance children’s sense of agency.
- Provide learning opportunities to sustain children’s motivation to learn.
- Support the development of skills at the children’s own pace as they progress through the game.
- Offer opportunities to engage in storytelling and elicit personal narratives.
- Provide opportunities to make contact with other patients and engage in synchronous communication.
- Support cooperation between learners.
- Reward children who seek challenges and take goal-attaining actions

![Figure 1. The framework of the Kala Forest DGBL system](image)
To enhance children’s interactions with peers, parents, and tutors, the multi-user technique was adopted. The Kala Forest system was developed using the Electro multi-user socket and designed using Adobe Flash and PHP, MySQL, and XML technologies. The system is composed of two subsystems: a narrative co-construction and sharing system and mathematics games. Based on the conceptual design, the system framework of the Kala Forest DGBL system is presented in Figure 1.

A “treasure box” panel is provided on the right side of the screen, where the user can choose an avatar (cartoon figure, e.g., Donald Duck, Superman, Panda) and username. The user can select the status of the room to indicate his/her availability for co-construction activities or to maintain privacy using the “stay open,” “agree to open,” or “closed” setting. When the room is open, other users may participate in construction by adding words or pictures to the hosts’ creations. For example, if a child’s room is open and the child writes a paragraph to describe the eating habits of his/her favourite animal, other children can insert words or sentences into the paragraph and co-construct the animal narrative together. Therefore, the narrative co-constructing and sharing system provides children with chronic illnesses the opportunity to narrate their life stories and share with other users as they wish.

Figure 2. (1) The main interface of the narrative sharing system and (2) the treasure box

Figure 3. The narrative co-construction and sharing system includes (1) graphic category icons, (2) graphic objects, (3) the typing area, (4) the sketching area, (5) drawing tools, and (6) the chat room
The mathematics subsystem of Kala Forest contains three games: Apple Catching, Shooting Balloons, and Whack a Mole (Figure 4). In these games, mathematics questions are randomly selected from an item bank drawn from a first–second-grade textbook and displayed. The user selects an answer to the first question to start the game. Every game action is associated with instant reward or feedback, with one “Kala dollar” awarded for each correct answer. To enhance the user’s motivation, s/he can play a Poke Fun game by trading 50 Kala dollars for a gift from the researchers. Images of these gifts appear in the user’s treasure box.

Figure 4. Mathematics games include (1) Apple Catching (2) Shooting Balloons (3) Whack a Mole

Procedure

John and Ted participated in the first iteration by using the Kala Forest DGBL system in weekly 2-h sessions for 5 weeks. The researchers and three graduate students served as adult tutors. Two adult tutors also bought their own notebooks and interacted with John and/or Ted face-to-face while they used the Kala Forest system. Our team guided the participants in drawing and sharing their experiences on different themes through the narrative co-construction and sharing system. For the mathematics games, the adult tutors assisted participants by explaining the questions, prompting responses, and providing feedback.

Findings

Five themes related to the children’s experiences were implemented with the narrative co-construction and sharing system in the first iteration (Table 2). Seven sessions of mathematics games were implemented. Due to differences in the participants’ medical treatments, the DGBL sessions were implemented individually, except in the fourth week, when John and Ted played together on the “motel design” theme in the narrative co-construction and sharing system.

The results of this iteration indicated that the narrative co-construction and sharing system stimulated children’s memories and motivated them to talk about their experiences with the tutors.

<table>
<thead>
<tr>
<th>Narrative co-construction and sharing themes</th>
<th>Mathematics games</th>
</tr>
</thead>
<tbody>
<tr>
<td>My favourite animals</td>
<td>Shooting Balloons</td>
</tr>
<tr>
<td>My favourite cartoons</td>
<td>Whack a Mole, Apple Catching</td>
</tr>
<tr>
<td>My favourite toys</td>
<td>Whack a Mole, Shooting Balloons, Poke Fun</td>
</tr>
<tr>
<td>My favourite gifts, Motel design</td>
<td>Shooting Balloons, Apple Catching</td>
</tr>
<tr>
<td>Places I hope to visit</td>
<td>Whack a Mole, Poke Fun</td>
</tr>
</tbody>
</table>

In the “my favourite animals” theme, John drew a pig and the tutor drew a lion (Figure 5-1). John talked about the drawing, saying:

Do you know that I can let the pig play on the slide? I and one of my friends played together with my pig toy on the stone slide at my kindergarten. I was very excited when I sat on the top of the slide and slid down.

In addition, the multi-user environment facilitated the children’s interactions via both face-to-face and online modes, and their engagement in diverse types of speech acts. In the fourth week, the tutors met Ted in John’s hospital ward. Both children were receiving intravenous therapy. They played with the narrative co-construction and sharing system on their own notebook computers side-by-side in the same bed. During the “motel design” activity, John said, “I am...”
waiting for you to login in the main hall... oh... I see your avatar.” John dragged and dropped graphic furniture objects into the co-construction sketching area. He said, “Ted, hurry up... you can choose the graphic objects on the left side of the screen.” Ted pointed the drawing and replied, “I’d like to put an airplane beside your furniture. I would like my mother to go with me.” John seemed very satisfied with their co-construction and said, “It is a wonderful motel that we can visit in the future.” (Figure 5-2). The above excerpt showed that children engaged in instructional, explanatory, and commentary speech acts, with acknowledging the intention and affect of the other player.

Figure 5. The sample artefacts on the narrative sharing system (1) my favourite animals (2) motel design

The interview data with his parents revealed that they had a positive attitude toward DGBL. The parents felt that the DGBL model exerted a good impact on the children’s learning in the hospital. They also provided some suggestions. Ted’s mother said:

*It is a very interesting computer programme. He can make drawings with new tools and have a fun time. The math games also improve his computational skills. I suggest that your team add more curriculum-related content to the programme rather than limited to mathematics. Because the learning opportunities for Ted are limited, I would like him to learn more.*

John’s father agreed this viewpoint, saying,

*It is too easy for him to play online games. The games with learning content are needed. I prefer to increase his learning abilities. Such learning experiences are helpful when he can re-enter school.*

John and Ted were strongly motivated to play the mathematics games at the beginning of the implementation period, but their motivation faded gradually. John had experience with large-scale commercial multi-user games. In the fifth week, he commented:

*You can develop many small games inside the Kala Forest. I like sports. Can you design skiing and bowling games? I have played caring-pet games before. You can design some pets for players to take care of. I had a cat in the game. I can give them milk, food, toys, and lots of other things online. When I take care of my cat, his life index gets higher. I think the caring-pet games are good idea for designing games.*

The tutor tried to encourage him by saying,

*You have a lot of Kala money. Let’s try it again.*

However, John lost his enthusiasm for continuing the Kala Forest mathematics games. Instead, he suggested that the tutors play new commercial games with him.

After the fifth week of the first iteration, John’s health status became critical. Research activities were thus suspended for 2 weeks, allowing us to redesign the model with consideration of the parents’ comments and children’s suggestions. We interviewed the bedside teacher about the perceptions of DGBL. The bedside teacher’s encouraging comments facilitated refinement of the DGBL model.

The bedside teacher, Miss Chen, had 9 years of teaching experience in the hospital. Observations of bedside teaching indicated that the participants anticipated Miss Chen’s class. Ted’s mother described the structure of Miss Chen’s instruction in Chinese language arts, including review, sentence composition, reading, and homework assignment. At
times, Miss Chen gave quizzes on mathematics or Chinese language arts. She noted that the DGBL model could enhance the children’s motivation, saying:

*Children are eager to learn the curriculum-related content. They like to treat themselves as normal students. DGBL can supplement bedside instruction that these children receive in the hospital or at home.*

The results of the first iteration indicated that the DGBL model required more curriculum content. Moreover, we integrated flexible and diverse features into the model to fulfill the needs of children with chronic illnesses. These processes yielded the multi-modal DGBL model proposed in this study. Multimodality is defined as the mixture of textual, audio, and visual modes in combination with media and materiality to create meaning (Kress, 2010). The multi-modal DGBL is defined as using the main interface of the game which integrates various small games for different interactions, such as “Who Wants to Be a Millionaire?” and Monopoly games.

**Second iteration: Implementation of multi-modal DGBL model**

In the second iteration, we examined the effectiveness of the multi-modal DGBL model for children with chronic illnesses in a hospital setting for ten weeks.

The process was the same as in the first iteration. John withdrew from the study due to his critical health status, and Ted and David participated in the second iteration. Their medical treatment plans differed and they moved back and forth between the hospital and their homes. Ted participated in 14 multi-modal DGBL sessions and David participated in eight sessions separately.

**The multi-modal digital game–based learning model**

We developed two versions of the “Monopoly-like” board game (basic and adventure versions) (Figures 6) that integrated the Kala Forest DGBL system with face-to-face educational activities. The Monopoly games used 3 × 3 grids, with each space representing a learning game. The basic Monopoly game consisted of eight activities in the following categories: computer-based Chinese language arts activities (sentence construction, auditory and written vocabulary exercises, multimedia story reading), face-to-face learning games (wood block stacking, origami, bingo), and digital games (a train game involving the ordering of numbers or vocabulary words, a basketball game in which the player shoots for the correct answers). The Adventure Monopoly game comprised eight activities, including four Kala Forest games and other computer games.

Players roll two dice and move clockwise on the basic Monopoly board. The six sides of each die represent one to four move steps and two chances for adventure. If a player rolls a “chance for adventure,” he/she switches from the basic to the adventure board game. The player has to roll the dice again. To enhance the children’s motivation, we changed activities every 2 weeks.

![Figure 6. The multi-modal board game (1) Basic version (2) Adventure version](image)
We redesigned the reward system. Players earned Kala dollars for each board game activity, which were collected on a reward card that listed gifts and their prices. A player could exchange the indicated number of Kala dollars for a gift, and could accumulate additional Kala dollars to exchange along with an already-obtained gift for a better, preferred gift.

Findings

Three features promoted the children’s motivation to participate DGBL activities. First, the multi-modal DGBL model integrated various activities that enriched the children’s learning experiences, thereby enhancing their motivation. Ted expressed great satisfaction with the various learning experiences:

I like to stack the wood blocks and play the basketball game. I never played with them before. I hope I can have an opportunity to play a real basketball game with my friends.

The simulated sports game appeared to prompt his wish to move beyond his physical constraints.

Second, the reward system used for the Monopoly board game significantly enhanced the children’s motivation to learn. The observation notes indicated that: After finishing the reading activity, the tutor asked David, “Which activity do you want to play next?” David answered,

I would like to move one step on the Monopoly board to play in the Kala Forest narrative sharing system and get 10 dollars. Or, I wish I can move 5 steps to play the train game and get 6 dollars. I need 5 more dollars to trade for a gift (lion mask) that I like.

David was very excited when he threw the dice and moved one step. At the end of the activity he expressed that he anticipated the tutor’s next visit.

Third, the refined rules of the Kala Forest mathematics games successfully sustained the children’s motivation. The games were redesigned as competitions between the tutor and participants to win the most Kala dollars. As David playing Shooting Balloons, he commented:

I feel that it’s not easy to use the space bar for shooting balloons. I saw the tutor playing very smoothly. Now I learned how to control my hands. I have to beat the tutor.

The multi-modal games also facilitated social interactions between the participants and tutors. During the sentence construction activity, Ted told a story about his dog. When asked to construct a sentence using the phrase “grow up,” he replied, I used to have a little dog. The dog grew up very quickly so that I could ride on it. It was my favourite pet One day, it got lost in the park.” During another session, David expressed his joy about leaving the hospital while making origami. The tutor observed that David drew smiley faces on two origami fishes. He said, “I can go home today. These fishes are celebrating for me.” The above excerpts showed that children felt free to share meaningful life events as well as project here-and-now feeling via the ongoing DGBL task.

The multi-modal DGBL model fulfilled the parents’ expectations for learning skills in the hospital. Ted’s mothers also commented:

The Monopoly board game facilitates Ted’s motivation to learn. He usually asks me how to win the games. I saw him always counting his Kala dollars to trading for a gift.

David’s mother reported:

The learning activities in the Monopoly board game expanded David’s learning experiences. David has never been to school. However, the revised games provide more “school-like” activities. When David goes to school, I believe that these experiences can help him adapt to school.
David recovered from his illness and attended school. The parents reported that the multi-modal DGBL experience contributed to his learning in school.

Discussion and conclusions

This study involved two iterations of design-based research to construct a DGBL model for children with chronic illnesses in the hospital environment. The first iteration revealed the need to enhance the children’s motivation and to provide more learning content, leading to the redesign of the model and successful implementation in the second iteration. The second iteration revealed that children with chronic illnesses showed more enthusiasm and preferred the activities included in the multi-modal DGBL model. Thus, this study demonstrated that the use of the multi-modal DGBL model was an effective and feasible educational strategy for children with chronic illnesses in hospital settings.

Children’s motivation plays a key role in the successful implementation of DGBL activities. Grossman (1975) indicated that depression and anger were the most common behavioural outcomes of hospitalisation in school-aged children. Participants’ emotions in our study were sometimes unstable due to medical treatment, which reduced their motivation to participate in the DGBL activities. These results echo the finding of Inal and Cagiltay (2007) that children’s flow experience, referring to a situation of complete absorption or engagement in an activity, is very limited. Csíkszentmihályi (1991) indicated that children’s flow lasted only 1-2 minutes and that they were distracted easily by other stimuli during game play. Therefore, the maintenance of chronically ill children’s motivation is an important issue in game design.

Game design should focus on motivational factors to promote a positive attitude toward educational games. In line with Malone and Lepper’s (1987) recommendations, fantasy, challenge, and rules were applied in this study. In the first iteration, fantasy was incorporated into the Kala Forest through avatar facility and the co-constructive narrative environment. In accordance with Dickey (2007), the character design and narrative environment of online role-playing may foster players’ intrinsic motivation and sustain their participation. The avatars available in Kala Forest allowed the children to represent their idealized self. We found that our participants selected the “Superman” avatar most frequently. Such a finding echoed that of Bers (2001) as her participants avoided mention of haemodialysis while in the virtual rooms. This was particularly true for ill children who were motivated to use avatars to escape from their suffering reality.

Challenge was implemented in the second iteration. Abiding by the participating children’s expectation, a composition of learning and commercial games was provided. In line with Facer et al. (2004) and Papastergiou (2009), a more sophisticated gaming environment was preferred by children. Moreover, Inal and Cagiltay (2007) indicated that children preferred switching among games when they played alone. The Kala Forest system provided only three mathematics games and thus few switching opportunities were available. In contrast, the incorporation of diverse games of the multi-modal DGBL system sustained the children’s motivation.

Moreover, we found that children with chronic illness were attracted by the simulated sports games in the second iteration. Due to their health condition, the children had been deprived of opportunities to engage in outdoor activities for a long time. Playing the sports games were novel experiences and thus challenging to these children. In addition, such games may relieve hospitalized children’s stress of a constrained body. As with Bers’ (2001) findings with Zora, the computer environment facilitated the patient to bring back their self-image as an active agent. Playing with the simulated sports games helped children move beyond the constraints in reality.

The game rules of DGBL promoted the children’s interest. The reward scheme redesigned in the second iteration intensified competition between the children and tutors. The awarding of points in an ascending scheme corresponding to levels of difficulty of the games facilitated the children’s engagement in challenging competitions with the tutors. Furthermore, according to Sutton-Smith (1997), the rhetoric of play as power views play as a representation of conflict and a way to establish and enforce the power status of winning players. The game rules that lead hospitalized children to resume a powerful identity are an important feature of the DGBL model.

In this study, a one-to-one tutor-pupil didactic strategy was used. In the hospital, time for play is unpredictable and not all children can engage in DGBL or participate in a virtual community at the same time due to the requirements.
of their medical treatments (Bers, 2001). In our study, only one opportunity for two participants to play together occurred, and the two boys did so with enthusiasm. Inal and Cagitay (2007) found that children were highly motivated when they played games with other children or adults. The results of the present study support the ability of side-by-side tutoring to fulfill children’s social needs. Wilson-Hyde (2009) indicated that consistent interaction with teachers is important for the educational progress and social development of young students with chronic illnesses.

The parents of participants in this study anticipated that the multi-modal DGBL model would have major learning outcomes. Our result was consistent with Lynch, Lewis and Murphy’s (1992) finding that parents with chronically ill children often express needs for tutoring services and home schooling to enhance children’s academic achievement. According to Ashton and Bailey (2004), many parents of children with chronic illnesses would like them to be treated as normal students and not differentiated from others, particularly when they return to school. Active participation in normal life experiences is one of the most effective ways to promote the mental health of children with chronic illnesses (Patterson & Geber, 1991). Activities encouraging children’s active engagement with others (e.g., peers, family members, and adults) promote the accomplishment of developmental tasks and contribute to psychological and social competence (Patterson & Geber, 1991). These experiences foster development by improving social skills, self-awareness, and identity. Future studies should examine ill children’s interactive experiences in larger communities of tutors and/or peers for longer periods of time.

The present study is inspiring to DGBL researchers. Future studies should explore the features of game design that incorporate multiple games into one large-scale game for promoting the motivation of children with chronic illnesses. Moreover, Kafai (1998) found significant differences in game design preferences between boys and girls. All participants in this study were boys. Gender-based preferences for game design should also be considered in future studies. Because few technological intervention studies targeting chronically ill children have been published, comparing our findings with those of other studies is limited. Future studies should incorporate large samples to obtain statistically significant results.

In sum, our findings suggest that DGBL provides new possibilities of learning continuity for children with chronic illnesses. The DGBL model in the current study incorporates the inspiring features including multi-modal activities, an ascending awarding scheme, and one-to-one tutor-pupil didactic strategy, which constitute a challenging game environment for ill children. Given the constrained experiences in daily lives, the hospitalized children not only earn learning experiences but also resume their interaction with social world. Future research should continue to explore the specific effects of the multi-modal DGBL model on children with chronically illnesses in diverse environments.

Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting under Contract No. NSC98-2511-S-152-016-MY2. The authors also want to thank the research assistants, Hao-Jan Luh, Hsin-Yin Chou, Hsueh-Huang Han and Levon Kwok, participated in this study.

References


The Nature of Online Charter Schools: Evolution and Emerging Concerns

Lisa Hasler Waters1*, Michael K. Barbour2 and Michael P. Menchaca3

1Technology Integration Specialist Flint Hill School, 10409 Academic Drive Oakton, VA 22124 // 2Farrington College of Education, Sacred Heart University, 5151 Park Avenue Fairfield, CT 06825 // 3Department of Learning Design and Technology, University of Hawaii, 1776 University Avenue Wist Hall Room 232, Honolulu, HI 96822

*Corresponding author

(Submitted October 01, 2013; Revised January 15, 2014; Accepted January 22, 2014)

ABSTRACT

Online charter schools are unique among K-12 online learning options for students. They are full-time, public schools that combine online learning with traditional and home schooling practices. They are often chartered by a state agency, supported in full or in part with state funds and most often managed by a private educational management company. Some extol the virtues of these schools as being able to reach unique student populations at a fraction of the cost borne by traditional public school education. Others are concerned over the lack of evidence supporting the effectiveness of these schools and the problems encountered by young learners who are separated from their teachers due to the online nature of learning in this environment. The goal of this literature review is to: (a) provide a definition of online charter schools; (b) describe their evolution and current status; (c) describe their operations; and, (d) to reveal emerging concerns, including governance, funding and effectiveness. Finally, the authors conclude that there are three significant gaps found in the literature concerning online charter schools and provide recommendations for further research.

Keywords

K-12 online, Online charter schooling

Introduction

Online charter schools are a cross between home schooling and charter schooling, in which technology plays a central role in the delivery and management of teaching and learning. Online charter schools offer an alternative to traditional brick-and-mortar schooling for elementary and secondary students and are unique among K-12 online learning options because they offer full-time online learning at no charge.

Historically, supplemental virtual schooling was the predominant form of online schooling in which K-12 students engaged. However, in recent years this has been shifting to favor more full-time online and blended learning options (Watson, Murin, Vashaw, Gemin, & Rapp, 2012). This virtual schooling might include credit recovery where students may need to make up for a failed course or opportunities for students to have access to advanced courses not available in the student’s regular brick-and-mortar school (Glass & Welner, 2011). Often, these types of virtual schooling options are offered to students through their own schools or school districts. However, full-time online charter schools differ because most can enroll students across district boundaries, and many are managed by private, for-profit companies that are contracted by a chartering agency, which might include state agencies, regional education services or a university.

Definitions

Importantly, not all K-12 online schooling is alike. The term K-12 online learning is generally used to refer to the practice of online learning for elementary and secondary students. Virtual school generally refers to supplemental programs that are offered online, and are taken by students who attend brick-and-mortar schools and who want to or need to supplement their course options. The term cyber school generally refers to a publicly funded, full-time school. Cyber schools are defined as schools that, “…work with students who are enrolled primarily (often only) in the online school” (Watson et al., 2012, p. 7). There are a number of terms that have been used to describe these specific cyber schools, including virtual charters, cyber charters, and hybrid charters (Glass & Welner, 2011; Huerta, Gonzáles, & d’Entremont, 2006; Klein, 2006; Rice, 2006; Vergari, 2009). Like brick-and-mortar charter schools, cyber charter schools are accountable to their granting agency, which in most cases is the state (Huerta et al., 2006).
Often, these schools are managed by a private entity or educational management organization (EMO) (Glass & Welner, 2011).

Some have also connected these cyber charter schools with the term blended learning, which has recently been defined as an education program in which a student learns “…in part at a supervised brick-and-mortar location away from home and at least in part through online delivery with some element of student control over time, place, path, and/or pace” (Horn & Staker, 2011, p. 3). Finally, Huerta et al. (2006) provided a definition of online charter schools that presents a more visual depiction—as non-classroom based charter schools because they deliver instruction beyond walls found in brick-and-mortar schools. For the purposes of this review, the authors use the term online charter school, which we define as a K-12 online learning program with the following characteristics:

- Publicly funded, usually by the state in which it operates or less often through public grants;
- Governed by the charter policies of the state in which it resides; and
- Relying on online learning and teaching for a significant portion of its delivery model, although it may combine other home and traditional schooling practices.

**Evolution and current status of K-12 online learning**

In the U.S., K-12 online learning in general has evolved from the long history of distance education. This history has paralleled many of the technologies used over the last two centuries—from print to media and communications technologies to the Internet revolution. Distance education is a “generic, all-inclusive term used to refer to the physical separation of teachers and learners” (Schlosser & Simonson, 2005, p. 84).

In the U.S., some of the first Internet courses began in 1986 as part of a program called the Quantum Link Community College project, which was located in New Hampshire (Darrow, 2010). The spread of personal computers also facilitated school use of computer-based and computer-aided instructional methods for supplemental practice and individualized instruction. According to Clark (2003) these tools, along with multimedia tools and creative interactive learning opportunities helped “set the stage for the virtual school movement” (p. 677).

With the Internet reaching beyond the walls of universities during the 1990s and expanding to the public, K-12 schools began to take advantage of this new medium for delivering education to younger learners (Clark, 2003). In these early years, much of the funding for K-12 online learning was supported by federal and state subsidies. One of the earliest examples of a school to provide supplemental virtual schooling was the Utah Electronic High School. According to Clark, this was followed by the Hawai‘i E-School, which was the first state-operated virtual school using only online instruction in the U.S. By 1997, the state-funded Florida Virtual School and the federally funded Virtual High School (later VHS Inc. and now VHS Collaborative) were both established. This latter initiative involved a consortium of high schools that crossed state, and eventual national, boundaries (Kozma et al., 2000).

Growth in K-12 online learning generally kept up a steady pace throughout the first decade of the 21st century. A group of researchers began tracking the steady growth in K-12 online learning across all 50 states and the District of Columbia, and learned that between 2004 to 2011 the number of states offering full-time and supplemental online learning programs for some students grew from just 11 states to all 50 plus the District of Columbia (Watson, et al., 2012; Watson, Winograd, & Kalmon, 2004).

Actual K-12 online learning enrollment numbers are somewhat difficult to come by because there currently is no single entity that tracks students, and because of the wide variety of ways in which students can engage in this form of schooling (Glass & Welner, 2011; Watson, et al., 2012). One report found that in 2010 there were over 1.8 million students enrolled in virtual schools (Queen & Lewis, 2011). This staggering number was dwarfed in comparison to reports emerging from the for-profit research firm, Ambient Insight (2011), which speculated that in 2010-11 over four million K-12 students in the U.S. participated in some form of online learning. Almost 300,000 of those students were enrolled in full-time online schools (Ambient Insight, 2011). Additionally, it was predicted that by 2016 there would be an estimated 4,750,000 full-time online K-12 students, and that 29% of all U.S. children would be enrolled in some type of online instruction.

Conversely, in its most recent report on the status of online schooling for the K-12 sector in the U.S., Evergreen Consulting predicted that growth in online charter schooling was slowing (Watson, et al., 2012). Evergreen
Consulting suggested that this could be attributed to the fact that full-time K-12 online learning required a time commitment that many parents might not be able to make. Horn and Staker (2011) asserted that this obstacle alone might prevent this type of schooling from growing beyond 10% of the total K-12 student population.

**Chronicling growth of online charter schools**

While online charter schools are a form of K-12 online learning, their operations are different enough to warrant a separate chronicling of their growth. From their slow birth in the mid-1990s, online charter schools grew significantly during the 2000s. This was perhaps due in part to the emergence of the for-profit companies, which develop online content for—and, in many instances, directly manage—many of the online charter schools today.

The growth of online charter schools has proved challenging to track. Figure 1 depicts the evolution of online charters from 1994 to a forecasted 2016. The figure is a compilation of a number of reports concerning online charter schools and illustrates the steady growth of this emerging form of K-12 distance education.

![Figure 1. Growth and evolution of U.S. online charter schools](image)

**Operations**

Online charter schools “come in many different flavors” (Ahn, 2011, p. 11). Some employ a hybrid model, in which students take their coursework online and are required to come to a resource or learning center at various times during the week. Others may enroll students across the entire state, so learning is conducted entirely online. These schools may rely extensively on parents or guardians (often referred to as learning coaches) to deliver instruction with the help of and assessment conducted by trained teachers.
Online charter schools also have unique attributes not typically found in traditional schools, such as flexible scheduling and the opportunity for students to learn at their own pace. These attributes are often what draw parents to enroll their students in online charters (Klein, 2006). Parents also enroll their students in these alternative schools because they offer increased learning opportunities, serve rural and otherwise isolated areas, offer flexible schedules to accommodate students who may be young professional actors or athletes, and they are convenient for students whose health may prevent them from traveling to and from a campus (Ahn, 2011; Erb, 2004). Other online charter students may come from at-risk backgrounds, while some may have special needs (Darrow, 2010; Hubbard & Mitchell, 2011a). Online charter schools also claim to facilitate learning at a pace suitable for the student, and offer the opportunity to catch up or get ahead academically because learning can be tailored to suit each child’s individual needs. Additionally, some parents choose online charter schools because they provide access to customizable education for free and because they align to parental values (Carr-Chellman, 2009; Erb, 2004).

Online charter schools share many commonalities with home schoolers. For instance, Bauman’s (2001) extensive study of 29,000 U.S. home schooled children found that many home schoolers were linked to online learning and online charter schools, thus supporting Carr-Chellman’s (2009) claim that there is an “inextricable link between home schooling and cyber charters” (p. 4). This could be in part because technology had facilitated learning at home (Andrade, 2008). It could also be perhaps because many home schooling students enroll in online charter schools to take advantage of the benefits provided by these public schools, such as the flexible schedules, quality of curriculum and freedom to learn in various locations (Huerta et al., 2006). However, unlike some home schooling, online charter schools are governed by state laws. Depending on the requirements of the state in which they operate, online charter schools may employ certified teachers and use standards-driven curriculum, and typically, students must take standardized tests (Revenaugh, 2005). Additionally, the parent is not in complete control of the curriculum as in a purely home school environment. For the parent or guardian whose child is enrolled in an online charter school, there is an added layer of accountability to the chartering agency that may not necessarily exist in purely home school environments.

There is rarely a typical day at an online charter school (Revenaugh, 2005). Flexibility is one of the unique features of these charters—where one student may prefer getting all his work done first thing in the morning, another may prefer working late into the night. Still, others may choose to work on math on Mondays and English on Tuesdays. The daily schedule is usually shaped by the parent and student, with input and guidance from the student’s assigned teacher. In Klein’s (2006) study on online charter schools she discovered that parents often had a difficult time describing a typical day. These parents reported that each day was shaped by the needs of the child and the tasks at hand.

While these schools are typically governed through some form of state regulation, many of them are directly managed by private organizations. Glass and Welner (2011) reported that over 75% of K-12 students enrolled in online charter schools that were managed by EMOs. These EMOs come in two forms: for-profit and not-for-profit. There were 33 states that allowed for-profit EMOs to operate schools in 2010-11 (Miron, Urschel, Yat-Augular, & Dailey, 2012). For example, the Open High School of Utah is an example of an online charter school that is not managed by an EMO (Barbour, Tonks, Weston, & Wiley, 2013). This school is chartered by the state of Utah and was established as part of a statewide policy initiative to allow broader access to educational options. It is unique in that it is relies on the use of open educational resources (OERs).

EMOs often employ administrative staff to manage the day-to-day operations of the schools they have been contracted to serve. They may also directly hire teachers to serve students. However, there are some cases where the school district contracting with the EMO will hire and manage certified public school teachers to serve students. Often, these EMOs receive between 5% to 40% less per student to manage the day-to-day operations compared to what their traditional counterparts receive from state funding agencies (Glass & Welner, 2011; Schaffhauser, 2012).

Online charter schools are unique, not only in form and function, but also in the elements that comprise the whole school. The technology, curriculum, students, teachers, and parents/guardians are each distinctive elements that function together in ways that are uncommon to their traditional school counterparts. In these schools, teachers use technology synchronously and asynchronously to deliver instruction to students and to communicate with them. For example, they may use an online presentation application, such as BlackBoard CollaborateTM, to conduct class or to engage in a collaborative project (Cavanaugh, 2008). They may work with a student individually using a chatroom,
or some other online interaction tool. Often, a third-party vendor online learning system (OLS) is used to monitor their students and to access content and instructional materials.

Some online charter schools offer a variety of curriculum choices, while others ascribe to just a single package (McCluskey, 2002). In the latter situation, a third-party commercial vendor typically provides curriculum and management (Huerta & Gonzáles, 2004), and usually outfits each student with curriculum-related materials (e.g., textbooks and manipulatives), a computer, and various hardware (e.g., a printer, headset, and microphone). Some schools may even reimburse families for a portion of their Internet subscriptions.

In most online charter schools, the student is assigned to a teacher, who may be certified, and who is often responsible for reporting student academic achievement, providing supplemental instruction, conducting assessments, and evaluating the student’s work and progress. Parents and/or guardians generally act as the learning coaches of their online charter students. As needed, they provide their students with instructional support, and the encourage them to make progress and help students to manage their learning time (Hasler-Waters, 2012). Some online charter schools provide a physical location where students can receive classroom-based instruction and engage in traditional school-like social activities, such as working in groups, doing community service projects, attending field trips and engaging in social events like holiday celebrations (Bogden, 2003; McCluskey, 2002). Others require that students attend at least one course or a full day at the school’s physical campus, while some may require more or less face-to-face time (Van Dusen, 2009). Further, some states allow online charter school students to participate in extra curricular activities at a regular brick-and-mortar school.

**Emerging concerns**

In the early 2000s, a handful of reports outlined concerns with the way these online charter schools were being managed and the lack of accountability required of these fledging schools. There was relatively little evidence that proved that these schools could achieve academic ratings similar to their traditional counterparts. This same apprehension still exists more than ten years after the emergence of these online charter schools. Results from empirical studies, state audits, investigative reports, and dissertations have presented concerning evidence that these schools are still troubled by (a) lack of oversight/accountability, (b) improper use of public funds, (c) failing grades, and (d) drop out rates that are higher than their traditional school counterparts.

**Accountability**

One of the key differences between online charter schools and brick-and-mortar charters is that online charter schools can serve students from across school district borders. This freedom from the “shackles of geographic boundaries” (Bogden, 2003, p. 33) challenges the governance of these schools because it makes accountability and funding problematic (Ahn, 2011; Barth, Hull, & St. Andrie, 2012; Carr-Chellman & Marsh, 2009; Schaffhauser, 2012; Vergari, 2009). There is concern over who is ultimately in charge of managing these schools, and whether or not public agencies have the ability to monitor the teaching and learning that takes place in a private residence (Barth et al., 2012; Huerta et al., 2006). Furthermore, there are very few agencies to help evaluate or provide reliable accreditation services for these relatively new forms of schooling (Glass & Welner, 2011).

Issues concerning the lack of oversight of and accountability for these schools have amplified as a result of some recent investigations and audits. In 2011, Colorado was cited as failing to provide proper oversight of, or sanctioning to, poor performing schools (Hubbard & Mitchell, 2011b); this was particularly noteworthy as the same criticisms were also found five year earlier (Office of the State Auditor, 2006). Arizona had yet to put into place independent oversight and measures requiring accurate reporting of student enrollment, completion rates and services for special needs students (Ryman & Kossan, 2011). Similarly, Minnesota was reported as not focusing adequate attention on student performance in these schools before re-approving certain poor performing online charter schools (Office of the Legislative Auditor, 2011). One investigative blogger was even able to uncover evidence that one online charter school had outsourced its grading to an overseas company in India, in clear violation of the state’s laws regarding personally identifiable information of students and fingerprinting of those involved in K-12 education (Safier, 2008).
The problems associated with accountability were coalesced into a report produced by the Center for Public Education (CPE), an initiative of the National School Boards Association. CPE researchers found that the lack of financial accountability stems from the fact that many of these schools do not have to release their budgets for public review as their traditional school counterparts must. Researchers called for “greater oversight and accountability to ensure virtual charter schools receive funding for those students they are actually educating” (Barth et al., 2012, p. 14).

Some state governments have addressed these concerns. For example, Wisconsin has put into place a well-defined authorization for online charter school operation and the creation of accountability measures (Stuiber, Strom-Horns, Kleidon, LaTarte, & Martin, 2010). Among its details, the legislation requires that for open enrollment to occur, the online charter school must be located within the school district that it had contracted with and that any person who teaches in a public school must be licensed by the state. Some statutory frameworks have been proposed to address systemic integrity by tying funding to performance and calling for more transparency (Bathon, 2011).

**Funding**

There has been extensive debate concerning how to fund online charter schools. Some question whether funding should follow the student or be distributed based on proportion of a district’s population. Others question whether start-up and maintenance costs are equal to the funding required for traditional schools (Barbour & Reeves, 2009; Carr-Chellman & Marsh, 2009). Some have campaigned against online charter schools claiming that their geographic freedom siphons funds from local schools and takes monies away from in-state departments of education by out-sourcing curriculum and management to out-of-state private vendors (Wisconsin Parents Association, 2002). At the center of the debate is the question whether these schools are more cost effective and should therefore receive less state funding to operate.

Some of the earliest cases concerning funding of online charter schools took place in Pennsylvania and California, and later in Arizona and Colorado. In Pennsylvania, funding for one online charter school, which served students from across the state and relied on school districts where its students resided to forward tuition payments, led to a fiscal crisis because schools refused to forward tuition to the online charter school. The issue was finally resolved when the state’s legislature passed Public School Act 88, which explicitly defined online charter schools as public schools and which required that they be granted charters only by the Pennsylvania Department of Education (2006). It also codified funding, stating that it was the responsibility of the student’s resident school district to make payments to an online charter school in which the student chose to enroll (Huerta & González, 2004).

In California, outrage ensued by those who alleged that private companies, which received public funds to manage online charters, reaped the benefits of receiving full funding for operating public schools with little facilities and smaller staff. This ultimately resulted in a drastic reduction of state funds allocated to these schools compared to their traditional school counterparts (Huerta et al., 2006). In a similar turn of events, the State of Arizona had problems tracking costs and allocating proper funding to its online charter schools because of problems associated with its electronic data recording (Ryman & Kossan, 2011). Additionally, a recent investigative report concerning Colorado’s online charter schools found that traditional schools were losing out on millions of dollars in student funding because they had to absorb those students who dropped out of the online schools, yet the funds to educate those students stayed with the online schools and the for-profit companies managing the students (Hubbard & Mitchell, 2011a).

Some have tried to determine whether these schools cost less to operate than traditional, brick-and-mortar schools and whether, as a result, they should be funded less than brick-and-mortar schools. The limited research that is available seems to suggest that these schools do operate more cost effectively, yet the answer is not quite as simple as the results imply. This uncertainty is primarily because data presented by these schools is selective (Barbour, 2013). Miron and Urschel (2012) asserted that in order to truly understand the actual cost associated with online charter schools, the corporations that operate them need to be “more transparent with their financial data than they have been heretofore” (p. 7).
Effectiveness

An ongoing question concerning online charter schooling is whether students in these environments achieve academically as well as their traditional school counterparts. A U.S. Department of Education report is often cited by proponents of online charter schools because it found students enrolled in courses that blended face-to-face instruction with online learning fared as well as their traditional school counterparts (Means, Toyama, Murphy, Bakia, & Jones, 2009). However, the authors of the study warned that the findings were focused solely on supplement virtual schooling and reflected results mainly analyzed from environments involving higher education rather than elementary or secondary schooling and thus did not fully represent outcomes related to full time online schooling for younger students.

Glass and Welner, (2011) produced a policy brief in part to analyze the political and economic forces shaping the growth and use of online learning in the U.S. and also to address concerns over the lack of empirical evidence demonstrating the academic effectiveness of these schools. Their report found that there was a severe lack of evidence demonstrating the effectiveness of student learning and achievement in full-time online schools. They warned that without such key information states should not look to expand full-time online charter schooling. A more recent report conducted by the Center for Public Education corroborated Glass and Welner’s (2011) concerns by recommending that since there was such a substantial lack of evidence supporting student achievement in supplemental and full-time online schools that legislatures needed to consider this before they expand online learning opportunities to K-12 students (Barth et al., 2012).

In a study comparing the achievement of students in non-classroom-based charter schools with their traditional school counterparts, Buddin and Zimmer (2005) discovered that non-classroom based charter schools in California had much lower test scores in reading and math than did their traditional school counterparts. In a subsequent study, Zimmer et al. (2009) expanded the previous study of charter schools to include seven additional states. While the study was on charter schools in general, one important finding concerned online charter schools was reported. When the authors considered achievement gains for students enrolling in Ohio’s charter schools, which entered students at Kindergarten, there was a significant and substantial negative gain. They attributed this to online charter schools because they constituted a large part of the enrollment of Kindergarten-entry charter schools in Ohio.

A more recent study conducted by Stanford University’s Center for Research on Educational Outcomes (CREDO) examined charter schools in Pennsylvania and discovered that all eight of the online charter schools included in the study performed significantly worse than their traditional school counterparts (CREDO, 2011). Further, four recent audits concerning online charter student achievement scores on state exams in Colorado, Wisconsin, Minnesota, and Arizona also showed less than favorable results. For example, a Colorado report that investigated reading scores of 2,729 full-time online charter students discovered that their scores dropped 6% over a one-year period (Hubbard & Mitchell, 2011b). Even more concerning, the report also found that students who switched from traditional schools to online charter schools saw their reading proficiency drop from 58% when they last attended the traditional school to 51% when they attended the online charter school.

Results from a State of Minnesota audit concerning K-12 online charter school students found that they scored comparable to their traditional school counterparts in reading, but lower in math (Office of the Legislative Auditor, 2011). The audit also found that between 2008-09 and 2009-10 full-time online students in fourth to eighth grade made only about half as much progress on the state’s standardized math tests as their traditional school counterparts. The students did, however, keep pace with them on reading tests. Similarly, auditors in Wisconsin discovered that the math scores of online school students were far lower than their traditional school counterparts but that their reading scores were comparable (Stuiber et al., 2010). A report by Arizona Republic, a daily newspaper serving Arizona, found problems similar to those encountered in Colorado, Wisconsin and Minnesota. Its online charter students were failing to perform at the same levels as their traditional school counterparts (Ryman & Kossan, 2011). The authors lamented that some of the larger state online school providers did not require in-person proctoring of final exams and reported that an audit in 2007 by the State of Arizona’s Auditor General's Office found that the state had no way to verify the number of hours students and parents reported doing course work and that it also could not determine whether the online courses improved student learning.

In the last quarter of 2011 a number of investigative articles in the popular press corroborated the findings of the state audits and questioned the quality of education in online charter schools. In particular, two investigative articles
produced by reputable, national newspapers showed that low student achievement scores were found in a number of online charter schools. A New York Times article contained findings concerning one online charter school, which enrolled students from across the state of Pennsylvania. The reporters found that 60% of its students were behind a grade in math and 50% in reading (Saul, 2011). Similarly, The Washington Post also conducted an investigation into the academic achievement of students who attended online charter schools in Ohio and Colorado. The reporters found that students in some of the schools had low on-time graduation rates compared to statewide statistics (Layton & Brown, 2011). One Colorado online charter school with enrollment of over 5,000 in 2010, faced an on-time graduation rate of just 12 percent versus the state’s traditional school students who achieved an on-time graduation rate of 72%. Likewise, several online charter schools in Ohio, with enrollments of over 9,000 students during the same period had a 30% on-time graduation rate compared to the statewide average of 78%.

Dropout rates

Compounding the issues concerning poor achievement are findings pointing to higher drop out rates among online charter high school students compared to traditional school. A Minnesota audit found that between 2006-07 and 2009-10 the drop out rates for its full-time online students increased (Office of the Legislative Auditor, 2011). While 18% of twelfth grade students dropped out during the 2006-07 school year, that percentage had grown to 25% of seniors during the 2009-10 school year. Moreover, the audit found that the traditional school seniors had only a 3% drop out rate at the end of the same period. Additionally, in Colorado, investigative reporters discovered that half of the state’s online students left their online schools within a year. These reporters discovered that when these students returned to the traditional school they were further behind academically then when they started (Barth et al., 2012; Hubbard & Mitchell, 2011a).

Similarly, Darrow (2010) found that California’s online charter students dropped out at higher rates compared to their traditional school counterparts. He discovered that over a two-year period, 2007-08 and 2008-09 the dropout percentage for students in online charter schools was between 22% and 59%, while the dropout percentage for students in the traditional schools ranged from 0.5% to 4%. Darrow posited, that while more data was needed to confirm his hypothesis, he believed that the high drop out rate could have been due in part because the online charter schools served a disproportionate number of at-risk students compared to the traditional schools. However, adding to the uncertainly concerning student characteristics and high drop out rates, Miron and Urshcel (2012) contended that online charter actually enroll a lower proportion of at-risk students that brick-and-mortar schools (Miron & Urschel, 2012). Clearly, more research must be conducted to better understand how to keep online charter students in school.

Conclusions

In order to better understand the role of online charter schools in a complex educational environment, this article reviewed the existing literature. In this section, the authors offer concluding statements summarizing the literature, describe three significant gaps found within the literature, and recommend future research. The authors provided a definition to clarify that these unique schools are publicly funded, governed by state laws, and rely on online learning for a significant portion of the delivery model. Their operations typically involve some mix of education management organizations, public school teachers and the parents/guardians of students. These schools expanded quickly in the early part of the decade. However, some believe their growth is slowing. The authors contend that emerging concerns (accountability, funding, effectiveness and dropout rates) suggest that there is great need to improve the quality of education and the effectiveness of operations in online charters.

The authors found three gaps within existing research. One of the more obvious problems is that most of the studies cover a relatively short period of time. The findings may in fact represent anomalies that may differ under longitudinal studies. Additionally, opposing views over the nature and composition of online charter school students suggest there are also deeper issues at play. Therefore, there is a need to learn why these students enroll in online charter schools in the first place. Beyond the student characteristics or demographic profiles, there could be environmental, social, and ecological factors that also need to be better understood.

Second, relatively few studies have covered familial issues and parental involvement in these schools. It has been suggested that some of these schools rely heavily on parents to support their children’s educational activities (i.e., as
learning coaches) and that these parents may not be fully prepared to take on such an instructional role (Hasler-Waters, 2012; Litke, 1998; Russell, 2004; Ryman & Kossan, 2011). Moreover, most parents are likely not certified teachers and may not be qualified to provide the type of educational support these students need (Ahn, 2011; Huerta & González, 2004; McCluskey, 2002; Stuiber et al., 2010). Additional research is needed on the type of instructional support online charter school students receive from their parents and guardians, as well as the teaching strategies employed in these schools as they relate to academic outcomes.

Third, much of the research conducted has only considered quantitative data on issues surrounding student performance and retention. Data of this nature are often not able to capture the complex implications associated with online charter schooling, such as the pedagogy, curriculum, or environmental and social factors that might influence students and their learning experiences. These more complex issues may be better understood through qualitative research techniques.

Future studies should also tend to the issue of fit, and whether or not certain students fare better in these alternative learning environments than others. Finally, relatively little research has captured the voices of students in these fully online charter schools. Having a greater understanding of student needs from their perspective may provide useful insight when building student support systems. In summary, more research involving the range of participants in these schools will lead to better implementation of online charter schools.

References


Collaborative Inquiry with a Web-Based Science Learning Environment: When Teachers Enact It Differently

Daner Sun*, Chee-Kit Looi and Wenting Xie

National Institute of Education, Nanyang Technological University, Singapore, 637616 // daner.sun@nie.edu.sg // cheekit.looi@nie.edu.sg // wenting.xie@nie.edu.sg

*Corresponding author

(Submitted October 17, 2013; Revised March 13, 2014; Accepted May 05, 2014)

ABSTRACT

Though discussion of the teacher factor in ICT-enabled science learning abounds in the literature, the investigation of Teacher Enactments (TEs) of ICT-facilitated lessons through exploring teaching practices is still under-explored and under-recognized. Current studies are still lacking in evidence-based findings of TEs based on the investigation of teaching practices. This study explores the TEs of science lessons supported by a web-based learning platform, namely, Collaborative Science Inquiry (CSI), by two experienced teachers in their respective classes. The CSI system is built on a model-based inquiry framework integrated with Computer-Supported Collaborative Learning (CSCL) elements. The CSI lessons selected were on the topic of “Diffusion and Osmosis” in Grade 7. Through examining the ways in which teachers instructed, questioned, and interacted with the students, we identified the commonalities and differences in TEs that subsequently influenced students’ conceptual understanding and their involvement in collaborative inquiry. The factors that contributed to these discrepancies were then discussed. Implications were proposed to inform the use of complex ICT tools in the science classroom.

Keywords

Collaborative inquiry, Science learning, Teacher enactment, Collaborative Science Inquiry learning environment

Introduction

With a burgeoning volume of work on the integration of Information and Communications Technology (ICT) tools into science education, much research has explored students’ learning processes and performances in the ICT-facilitated science classroom (Buckley et al., 2004; Jacobson & Archodidou, 2000). Recently, research focus has been placed on investigating the relationship between teacher attributes (e.g., technological pedagogical content knowledge, beliefs and attitudes) and their teaching practices with the intention to improve the quality of ICT-facilitated instruction (Song & Looi, 2012; Voogt et al., 2013). Despite the rapid technological advances, pedagogically effective and sustainable use of ICT tools in education is still far from reality (Dimitriadis, 2010). In the ICT-facilitated classroom, teachers need to coordinate ICT and non-ICT activities and artifacts, and handle different levels of social interactions (Dillenbourg, Järvelä, & Fischer, 2009). Hence, Teacher Enactment (TE) of the ICT-facilitated lessons in the classrooms has been regarded as a critical indicator for evaluating teacher performance on ICT integration in the classroom.

An in-depth understanding of teacher practices is indeed crucial for the successful implementation of ICT-based initiatives, yet most studies only employed self-reported teaching beliefs and practices as the sources of evidence to probe these issues (Mama & Hennessy, 2013). The dominance of and constraints inherent in survey studies motivate an in-depth investigation of TE of ICT-facilitated science lessons by directly observing the actual happenings in the classroom. This study attempts to represent, interpret and compare the TEs by two experienced teachers of a lesson design that incorporated the Collaborative Science Inquiry (CSI) learning environment. Through fine-grained analysis of classroom practices, the commonalities and differences in TEs and student performances were identified. The findings can inform the effective use of ICT tools in science instruction via bridging the gap between the intended lessons and their actual enactment.

An overview of the CSI learning environment

Incorporating multiple features that support collaborative science inquiry, CSI is designed to help secondary school students (Grade 7 - 11) develop sophisticated understanding of scientific concepts, scientific skills (i.e., modeling
skills) and reflective thinking skills (Sun & Looi, 2013). CSI consists of two functional modules: Teacher Module and Student Module. Teacher Module encompasses four sections including Subject Management, Project Management, Simulation Library, and Solutions Review, with which teachers can design instructions and questions, attach simulations, manage groups, and review learning artifacts. Figure 1 illustrates the interface of “Project Management” used to create and manage inquiry-based projects. The inquiry phases of Overview, Contextualize, Question & Hypothesize (Q&H), Pre-Model, Investigate, Model, Reflect and Apply are incorporated. Teachers can select an intended inquiry phase, instantiate it with specific content and instructions, and employ modeling or visualization tools depending on the lesson objectives.

![Figure 1. Interface of “Project Management”](image1)

Student Module is comprised of Profile, My Project, Group Management, and Mailbox. The main tool “My Project” consists of four panes: inquiry phases, shared workspace, group members, and chat tool. Student inquiry is guided by the inquiry phases laid out on the tool bar (Figure 2). The shared workspace stores the content and tools for completing tasks at each phase. The embedded Computer-Supported Collaborative Learning (CSCL) design elements (e.g., shared workspace, chat tool, peer review, and social presence) enable students to do various forms of collaboration during the inquiry activities.

![Figure 2. Interface of “My Project”](image2)
Theoretical framework

Pedagogical model of CSI learning environment

The design of CSI is motivated by the educational benefits brought about by model-based science inquiry and CSCL. Science inquiry is a process where students ask questions, search for information, design and conduct investigations, analyze data and make conclusions, create artifacts, and communicate findings (Krajcik et al., 1998; NRC, 1996). When integrated with models or modeling, it can facilitate learners to develop deep scientific understanding, strong scientific skills and a solid understanding of the nature of science (Schwarz & Gwekwerer, 2007). Model-based inquiry that follows the sequential processes of question-hypothesis-plan-investigation-model-conclusion proves to be effective for science learning (Bell et al., 2010; Schwarz & White, 2005; Windschitl, Thompson, & Braaten, 2008) and is adopted in the design of CSI. Informed by the POE (Predict-Observe-Explain) instructional principle practiced in some Singapore schools (White & Gunstone, 1992), a Pre-Model phase that precedes the Model phase in inquiry is incorporated. The main purpose of embedding model progression (from Pre-Model to Model) into the science inquiry is to guide students to elicit prior knowledge through constructing pre-models before investigation and to elaborate models after inquiry activities. Finally, the model-based inquiry of CSI incorporates eight phases that run in sequence, namely, Contextualize, Question & Hypothesize (Q&H), Pre-Model, Plan, Investigate, Model, Reflect, and Apply.

Besides the presence of a Pre-Model phase in inquiry, the uniqueness of CSI also lies in the integration of CSCL elements. Informed by established design principles and applications (e.g., WISE, CMapTools, Co-Lab, and ModelingSpace) that incorporate CSCL ingredients to further empower learning (Avouris et al., 2005; Cañas, Novak, & González, 2004; Linn, Clark, & Slotta, 2003; van Joolingen et al., 2005), synchronous modeling and editing, shared workspace, peer review, chat tool, and social presence are employed in the CSI inquiry. In Overview and Contextualize, online members can view the problem statement presented. Students work in small groups to respond to the inquiry questions and problems. Students are allowed to edit and revise their answers synchronously. The shared and synchronized workspace in Pre-Model and Model allows for inputs from multiple devices to support concurrent multi-user operations, such as co-constructing, reviewing or revising models in real time. The design is intended to encourage students to pursue the common goal of creating joint models through processes of collaboration and interaction. Coupled with a chat tool, each inquiry phase supports synchronized peer discussion. Thus, the unique feature of the CSI system is the marriage of relevant CSCL functionalities with each inquiry phase, such that each phase can be utilised in a flexible way towards collaborative inquiry learning.

ICT integration into teaching and learning

The way in which technology actually appointed in a classroom for teaching and learning is a critical measurement of its success. Ertmer (2012) distinguished two types of barriers that hinder ICT integration. First-order barriers include resources, training and support, and second-order barriers include teacher confidence, beliefs and their perception of the technology. With advances in ICT and support from policymakers and administrators, the first-order barriers are gradually lessened, and now the focus is to address the second-order of barriers. Existing research has discussed teacher perception of the technology and their pedagogical approaches to ICT integration intensively. Baylor and Ritchie (2002) advocated the use of ICT as cognitive and integral tools in the curriculum to support the development of existing cognitive structures and new knowledge in students. Moreover, constructivist approaches of instruction, such as conducting learner-centred activities, asking exploratory questions, and providing flexible scaffolds are advocated in ICT-facilitated lessons (Hermans et al., 2008). Through these strategies, students will become active learners who benefit more in both knowledge and skills development.

The design features of CSI are intended to guide the teachers to integrate ICT tools in a more constructivist way. The CSI inquiry encourages students to pose hypothesis, investigate scientific phenomena, construct scientific models, collect evidence and reflect upon the processes in and out of classroom. This may enhance learner autonomy in learning. It offers opportunities for students to discuss solutions, co-construct knowledge, assess artifacts, and interact with teachers. With frequent use of CSI, teachers’ traditional pedagogical approach to ICT integration will be shifted to the constructivist approach.

Teacher enactment in collaborative inquiry learning environment

To explore how the teachers use CSI in the classroom, we identified some indicators based on a literature review. Previous research showed that good TE could not be achieved without appropriate facilitation. Crawford (2000)
found that teacher responses to the key instructional events and their roles acted in the inquiry phases were the key factors for inquiry-based instruction. He suggested that teachers should play a wider range of roles in facilitating inquiry activities. Abdu et al., (2012) advocated that teachers should serve as moderators to provide three types of assistances (i.e., presenting challenges, supporting group collaboration, and scaffolding meaning making) to establish successful teacher-student interactions. Onrubia and Engel (2012) pointed out that patterns of teacher assistance with the use of macro-scripts affected students’ plans, organization and coordination of work, and levels of achievement in collaboration. Chiu (2004) found that teachers’ initiating interventions, evaluating student’s work, and using higher levels of cognitive assistances could generate positive educational effects. From these studies, it is known that the ways teachers assist, respond to, and intervene in students’ work are important indicators for evaluating TE, and thus they are used in our study.

These studies also show that patterns of teacher-students discursive interactions reflect the nature of classroom activity (Jones & Charles, 2007). Three categories of teacher verbal behavior are identified, namely, instruction, question, and scaffoldings for mediated-learning (Gillies, 2006). Instructions are lectures that provide content facts and explanations, giving expression onto teacher ideas and opinions (Flanders, 1970). Questions are enquiries of content or procedures with the intent that students can answer. Mediated-learning is a way of interaction between the teacher and students. As a mediator, the teacher may provide different scaffolds through the form of micro- or macro-scripts to regulate the processes through the sequencing and distribution of roles and activities (Weinberger & Fischer, 2006). Teachers may also provide hints, suggestions, and reminders in the form of prompts to help students complete tasks (Ge & Land, 2004; Morris et al., 2010). Moreover, teachers can scaffold potential learning through challenging students’ thinking and encouraging them to consider alternative perspectives. When teachers frequently encourage autonomous behavior by students, students show more initiative and willingness to explore activities (Prieto et al., 2011). These factors influence the completion and quality of student work in ICT-facilitated lessons.

Research questions

The goal of this study is to generate teaching strategies that can improve the enactment of CSI lessons via characterizing and comparing TEs by different teachers and exploring the impact on students’ science learning. To achieve this goal, the following research questions will be addressed:

- What were the major differences between the desired TEs as proposed in the lesson design and the actual TEs as observed?
- What were the major differences in TEs by two different teachers when they implemented CSI lessons?
- How did different TEs affect students’ performance in collaborative inquiry?

Methods

Participants

In this study, two science teachers - Katherine (Class K, n = 21) and Charley (Class C, n = 20) (pseudonyms) and their respective classes from a junior secondary school in Singapore were selected as the participants. Katherine and Charley were comparable in their ages, length of teaching experiences, and educational backgrounds. Both possessed good ICT skills and used ICT as their instructional tools in the classroom. Through attending regular CSI project meetings, the two teachers had gained some understanding of the system design and its underlying pedagogy. Both teachers had strong enthusiasm in transforming their pedagogical orientation from the traditional way to the constructivist way. In the collaborating school, each student owned and used a MacBook for his or her daily lessons. In CSI lessons, students mostly worked in pairs using their personal laptops.

CSI lesson design

The topic of “Diffusion and Osmosis” was identified as one of the most difficult topics in Grade 7 science curriculum (Odom & Kelly, 2001), and was thus selected for implementation and analysis. The researchers and teachers co-designed CSI lessons consisting of two consecutive lessons (50 minutes per lesson) which followed the order of (Overview) → Contextualize → Q&H → Pre-Model → Investigate → Reflect → Apply (Table 1).
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Proposed Teaching Strategies for TEs</th>
<th>Form of Activity</th>
</tr>
</thead>
</table>
| Overview | • Introduce learning objectives  
• Emphasize tasks in different inquiry phases  
• Remind students to click task checklist when they have completed their work  | Individual |
| Contextualize | • Present and extract the key information  
• Pose guiding questions  | Individual |
| Q&H | • Encourage peer discussion  
• Assist in and coordinate students’ synchronous writing  
• Review students’ collaborative work and provide assistance  | Collaborative |
| Pre-Model | • Ask students to review “Instruction”  
• Encourage and assist in students’ individual modeling activities  
• Observe and assist in peer review and peer discussion of individual models  
• Encourage and assist in peer discussion and peer work in building models together  
• Observe, review and assist in collaborative modeling activities  
• Present students’ typical models and highlight misconceptions  | Individual and collaborative |
| Investigate | • Ask students to manipulate and observe simulations individually  
• Encourage and assist in peer discussion and collaborative answering of guiding questions  
• Encourage and assist in students’ collaborative work  | Individual and collaborative |
| Reflect | • Emphasize critical reflection on work produced in Pre-Model and Q&H  
• Encourage students to reflect upon their process of conceptual changes, if any  | Individual and collaborative |
| Apply | • Emphasize and assist in individual work  | Individual |

In the first lesson, students reviewed the textual information in Overview. In Contextualize, a story was introduced to arouse students’ interests and motivation. In Q&H, students discussed and articulated their responses to two inquiry questions posed. In Pre-Model, students watched two videos of lab experiments (the diffusion of red ink in water and the changes of raw egg in corn syrup and in water) to gain some ideas on the macro-phenomena of diffusion and osmosis. Students then were required to build models to represent the processes of diffusion and osmosis at the particulate level in the individual modeling space, and to collaborate with their partner to elaborate their shared models in the group modeling space (Figure 3). In the second lesson, the teacher summarized students’ work in Pre-Model and Q&H, and selected some work for plenary class sharing. Then students continued their inquiry, interacting with three simulations and answering the questions based on their observations of the virtual experiments in Investigate. After this, each student did a self-reflection on their conceptual change and learning process. Finally, students consolidated their new understanding via answering questions in Apply.

![Figure 3. Interface of pre-model](image-url)
Data sources and analysis

To answer the research questions, a case study approach was adopted. In analysis, the first focus was on teacher verbal interaction and assistance to specific students or groups. The data collected include videos and audios of classroom activities (e.g., inquiry activities, group activities, and individual activities) and field observation notes (where teacher questions, scaffoldings, and responses to students were recorded) taken by two researchers. Based on these data, the types and frequency of teacher verbal behavior relative to instructions, questions, scripts, prompts and challenging students’ ideas in key instructional events (e.g., Q&H, Pre-Model, Investigate, etc.) were identified and analyzed. The recipients (i.e., the verbal behavior as to who it was targeted at, namely, to the individual or group, or class) were also coded (Prieto et al., 2011). To identify the differences and commonalities in the TEs, a diagram was constructed to represent teacher verbal behaviors, such as how they performed in instruction, how they scaffolded students’ group work, and what kind of scaffoldings they offered at the different phases of inquiry. To identify the roles teachers played in the CSI lessons, the patterns of teacher facilitation (e.g., frequency, content, and recipients) were also examined at each phase (Onrubia & Engel, 2012).

To explore the impact of TE on student learning, data on student test scores, learning artifacts in CSI and their performance in collaborative work were collected and investigated. A pre-test and post-test (10 minutes for each test) using identical test items were adopted to probe students’ conceptual change. In the test, 10 paired questions were designed based on the validated two-tier “Diffusion and Osmosis Diagnostic Test” (DODT) (Odom & Barrow, 1995) (The tests can be retrieved from: https://sites.google.com/site/futureschoolcsinquiry/pedagogical-resources/diffusion_osmosis_test). Each correct answer was assigned 1 mark, so the total score was 20 marks. A paired-samples t-test was conducted to examine students’ conceptual change in each class. Student learning artifacts created in Q&H, Investigate and Apply, individual and collaborative pre-models and reflections were mined and assessed as to identify their completion rates and quality levels. Finally, students’ involvement in collaborative inquiry tasks and peer discussion were examined. Figure 4 shows the structure of the data sources and analysis.

The transcription and analysis of the qualitative data was conducted by the two researchers. The inter-rater agreement reached 89.15% for teacher verbal behavior, 92% for patterns of teacher facilitation, and 95% for students’ learning artifacts, and 93.46% for student performance.

Findings and discussions

Teacher verbal behavior (VB)

Table 2 shows Charley’s and Katherine’s verbal behaviors during their TE. As we observed, Charley acted as a guide and mentor who offered instructions for specific tasks before the activity. Being not frequently involved in students’ peer discussion, Charley spent most time in prescribing scripts and walking around the class to check and monitor students’ progress. Consequently, 6 instructions and 18 scripts were delivered. Katherine was involved in peer discussion in most groups, explaining the tasks to the students. More prompts (38) were generated in her lessons.
guide students’ conceptual understanding, Katherine often challenged students’ existing ideas through questioning (3). Thus, in CSI lessons, Katherine acted as a motivator, diagnostician and collaborator in students’ collaborative inquiry.

<table>
<thead>
<tr>
<th>Categories of VB</th>
<th>Instructions</th>
<th>Questions</th>
<th>Mediated-learning</th>
<th>Scripts</th>
<th>Prompts</th>
<th>Challenging ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charley (Class C)</td>
<td>6</td>
<td>3</td>
<td>18</td>
<td>16</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Katherine (Class K)</td>
<td>3</td>
<td>3</td>
<td>15</td>
<td>38</td>
<td>13</td>
<td>0</td>
</tr>
</tbody>
</table>

Charley’s verbal behavior

As Figure 5 shows, all the verbal behavior observed in Charley’s lessons was targeted at the class level. This indicates that he laid great emphasis on classroom management through offering task-related scripts and procedural prompts (9 instructions and 34 mediated learning scaffoldings with 18 scripts and 16 prompts). He tended to manage the class in the stages of pre-test, system login, overview and inquiry (i.e., Q&H, Pre-Model, and Investigate). The regulation of group and individual work was neglected as there was little talk directed to individuals or groups. Before the inquiry, Charley went through all the tabs by introducing the purpose, procedures, and sequence of the tasks at each phase. Strategies on the use of chat tool, peer review and assessment, and collaborative modeling were also elaborated. During inquiry, he spent most of the time providing methods and procedures for task completion in a step by step manner, such as:

Charley: Now I have two persons, later you will see your name there and your partner’s name there. For example, you can key in one sentence. Then your friend thinks that “I can improve your sentence”. So “my friend said that ...” (types on the system) and so you can continue, are you clear?
Charley: So to prevent overwrite of your answers, there is one way. You and your partner try to strategize it. You key in the answer to the first question, and then your friend keys in the answer to the second question. When you have finished, you tell each other you have finished, and then you go and check, and edit (the answer to the other question). Then you won’t overwrite each other’s answers.

Although most students managed to follow the instructions and complete the activities assisted by the scripts and prompts provided, few scaffolds were offered to groups and individuals. In several groups, some low ability students were unclear of what they were expected to do in the tasks. Even though they raised their hands for help, Charley did not attend to these requests and continued to move rounds among groups as he focused more on the progress of the whole class. In addition, his strict management of time made these students nervous. Peer discussion was also interrupted. This led to the emergence of passive attitudes, relatively low work completion rate, and infrequent peer interaction in some groups.

Katherine’s verbal behavior

Unlike Charley, Katherine was always busy walking through the groups, assisting groups or individuals at each inquiry phase. As Figure 5 shows, a similar frequency of scripts and prompts was observed at each phase. This demonstrates that Katherine had competence in managing the students’ progress. She was more frequently involved in students’ collaborative activities and discussion in inquiry, especially in the Pre-Model phase. 8 prompts for groups and 10 prompts for individuals were found at this phase. She provided immediate feedback to students’ requests and acted as an adoptive facilitator to help students complete the tasks at both individual and group levels. Compared with Charley, Katherine’s instructions and mediated-learning scaffoldings were more related to the content knowledge of diffusion and osmosis. She was good at motivating groups’ deep thinking of concepts through challenging established ideas or knowledge. Though the purposes and objectives of each phase lacked elaboration, Katherine highlighted the sequence of tasks at each phase, and steered students’ activities and collaboration towards the right direction. Generally, Katherine’s instruction was more student-centred. She created a comfortable discourse environment where she pointed out students’ misconceptions, described how she expected the students to converse, listened to students’ ideas, clarified contributions, and provided suggestions (van Zee et al., 2001).

More specifically, during her interactions with individuals or groups, Katherine focused on reviewing and commenting on students’ work. Katherine was able to comment on students’ current understanding and discussion. In particular, she guided students’ modeling tasks through reasoning the components of diffusion and osmosis,
explaining the possible changes of particles (i.e., states, sizes, and numbers) before and after diffusion, challenging students’ previous ideas, and leading them to gradually comprehend the new knowledge, such as:

Katherine: You have learnt molecules right? In chemistry, what are water molecules? Will they be exactly the same as other water molecules? If they are all water, when you draw the H₂O right, (the water molecules) should be the same, is it? Is it possible that I have another water molecule that is bigger with a lot of other atoms or things, or do they actually have the same number of atoms, and the same size of the molecule?

Encouraging students to analyse their own thinking and misconceptions helped them to predict, identify, and generate solutions. In comparison with Charley, Katherine provided her students with limited structural and procedural information before the tasks. Consequently, the students generally lacked understanding of the task purpose. Some students were confused, unfocused and unproductive in their work. This resulted in a considerable amount of requests for clarification on the task procedures and purposes from the students. So, the time management issue arose for Katherine as more class time was spent in answering students’ questions or requests.

Teacher facilitation

In CSI lessons, teachers tended to provide students with appropriate help for the completion of the tasks, the coordination of collaborative work, and the understanding of intended concepts. Figure 6 depicts the frequency of teacher assistance at each inquiry phase. From pre-test to Investigate, substantial assistance was offered by both teachers (34 times in Charley’s class; 70 times in Katherine’s class). Katherine provided more of these as she actively diagnosed and facilitated students’ problems. Charley provided general assistance to the class when he reviewed students’ work. As Figure 6 shows, Charley provided more assistance at the beginning phrases (from pre-test to Q&H). Starting from Pre-Model, Katherine provided more assistance than Charley, especially in individual and group activities. An explanation for the dramatic reduction of assistance from Charley might be students’ focusing on observing simulation and doing self-reflections and thus requiring limited structural information. With regard to the recipients, Charley’s assistance was primarily targeted at the class (24 times for class, 11 times for individuals) before and during tasks (17 times before, 15 times during, and 2 times after tasks). Katherine preferred to offer assistance to groups or individuals (25 times for class; 45 times for individuals) while they were doing their tasks (11 times before, and 59 times during tasks).
Students performance

Test achievements

Results of paired-samples t-test showed that students’ test scores of the pre- and post- tests differed in both classes (Class C: \( t = -4.152, \ df = 16, \ p = 0.001 < .05 \); \( t = -5.920, \ df = 18, \ p = 0.000 < .05 \); Class K: \( t = -5.920, \ df = 18, \ p = 0.000 < .05 \) (Table 3). Both classes had improved their test scores after the CSI lessons. Class C had comparatively better prior knowledge \( (M = 10.53; \ SD = 2.503) \) than Class K did \( (M = 8.53; \ SD = 2.695) \). Yet the disparity of mean scores between Class C and Class K was reduced from 2 to 0.97 after the CSI lessons.

Table 3. The results of paired samples t test

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pretest M</th>
<th>SD</th>
<th>Posttest M</th>
<th>SD</th>
<th>n</th>
<th>Difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tc/Class C</td>
<td>10.53</td>
<td>2.503</td>
<td>14.18</td>
<td>3.187</td>
<td>17</td>
<td>-5.509</td>
<td>-1.785</td>
<td>-4.152</td>
<td>16</td>
</tr>
<tr>
<td>Tk/Class K</td>
<td>8.53</td>
<td>2.695</td>
<td>13.21</td>
<td>2.573</td>
<td>19</td>
<td>-6.347</td>
<td>-3.022</td>
<td>-5.920</td>
<td>18</td>
</tr>
</tbody>
</table>

The results indicate that the CSI lessons in general could improve students’ conceptual understanding. The class (Class K) who had less prior knowledge benefited more. It is inferred that:

- The CSI lessons could enhance students’ conceptual understanding through collaborative inquiry approach;
- The guided-inquiry and collaborative work in CSI lessons facilitated low ability students in learning abstract concepts;
- TEs may affect students’ conceptual understanding.

The teacher who paid more attention to diagnosing students’ misconceptions and providing assistance for conceptual understanding might be more effective in enhancing students’ conceptual understanding.

Learning artifacts

Both classes went through the designed tasks, but their task completion degree and work quality differed. This could be ascribed to the discrepancies of TEs. As aforementioned, Charley provided more instructions at the beginning phases of inquiry. Consequently, students’ performance differed in the Q&H phase. More students in Class C completed their work (73.5%) and responded appropriately to Q1 (53.4%) and Q2 (44.4%). In Class K, 60% students responded in the Q&H phase, with 50% and 35.3% appropriate answers to Q1 and Q2. This indicated that the elaboration upon skills for peer review and discussion by Charley facilitated students’ responses to questions at Q&H. For Class K, students’ confusion about the purposes and skills for the Q&H tasks resulted in students’ failure to complete the task.

In Pre-Model, Class C and Class K constructed comparable numbers of individual models (Class C: \( n = 16 \), Class K: \( n = 14 \)). With regard to model quality, Class K performed better. For Class K, 10 models (out of 14) represented the components of diffusion and osmosis and their relationship links (though sometimes the relationships were partially incorrect), yet for Class C 8 models (out of 16) contained incorrect components and relationship links. This indicates that greater involvement in peer discussion and more scaffolds provided for modeling by the teacher facilitated students’ understanding of relevant concepts (i.e., the components of diffusion, the process of diffusion, and the mechanism of osmosis). For collaborative models, students’ work completion degree was not as high as the individual ones. This could be resulted from the limited class time.

Similar findings were found in Investigate. Both classes provided answers after they observed and manipulated simulations, and their knowledge about diffusion and osmosis was improved. Although Class K received lower scores in the pre-test, they responded equally well to the questions in Investigate as Class C did. For some questions, their answers were even better. In Reflect, Class C provided more critical reflections (45%) than Class K did (36.4%). Class C went through the dual processes of self- and peer-reflection directed by Charley, yet Class K solely focused on self-reflection due to the lack of emphasis and guidance. In Apply, Class C and Class K both provided highly correct answers to the three questions (83% on average). This further indicates students’ improvement in conceptual understanding as they could apply the knowledge to new contexts.
Performance in collaborative tasks

Class C

In Class C, the TE of tasks was consistent with our lesson plan. Students had clear understanding of the tasks. In the collaborative work, students sat normally and chatted with their group members using the chat tool. The classroom was quiet and in order (Figure 7a). “Noise” was only heard during the Pre-Model phase as students in the same group sat together to do face-to-face discussion. At this phase, students shared ideas with their group members to review, evaluate and elaborate their models (Figure 7b). Most groups were highly engaged in open and extended discussions. As they received substantive scripts and prompts from the teacher, they rarely asked for assistance. Below is an excerpt of a selected student dialogue from Class C:

[S1]: OK, you draw diffusion and I draw osmosis.
[S2]: Huh? This is not I do osmosis.
[S1]: Yes, which side do you prefer?
[S2]: I will take right side.
[S1]: Ok. I draw at left side. Let’s do it.

Figure 7a. Students were quiet during Q&H phase

Figure 7b. Students were talking during Pre-Model phase

Class K

In Class K, students sat with their partners throughout all collaborative activities as proposed in the lesson plan. Most students chatted with their partners in a face-to-face manner. Hence, more “noise” was heard (Figure 8a). There were multiple instances in which Katherine provided assistance to individuals and groups. Students engaged in active interaction, sharing and discussing ideas. Their collaboration was more productive in terms of knowledge understanding. They had strong willingness to invite Katherine to join their discussion to evaluate their artifacts, share understanding of concepts, and explain task purposes and procedures (Figure 8b). However, with limited knowledge and skills on collaborative learning, students requested more scripts regarding labor division and task procedures from the teacher. Here is an excerpt of a student dialogue from Class K:

[S1]: I had no idea what is the diffusion.
[S2]: That’s what I’m typing now. Diffusion is molecules spreading throughout the air from higher concentration to lower concentration.
[S1]: Higher concentration to a lower concentration. Ethan was able to smell the fishes, the cooked fishes due to diffusion of molecules ...

In Pre-Model, one student requested Katherine to review and evaluate his model quality. He responded to the teacher’s questions and followed her instruction on the modeling.

[S]: Can you help me look through this drawing?
[Katherine]: But they are all different sizes. Shall all ink particles be of same sizes? They are of the same or different particles?
[S]: They are all the same.
[Katherine]: Did you copy and paste (the particle drawing)? First copy, paste (the particle representation) and get them all of the same size.
[Katherine]: Why then have one (particle), is that clear (some student name)?
[S]: To get the same size. That's the main aim.
[Katherine]: So you draw a lot of rings. A lot of the same size, but all concentrated in that area. That should be pretty sufficient already to represent.

Figure 8a. Students were working together in collaborative tasks

Figure 8b. Katherine was involved in students’ discussion

Conclusions

Much research has identified that teacher attitudes and beliefs toward technology and their technological knowledge and skill predict their technology use (Ertmer, 2012; Liu, 2011; Tondeur et al., 2012). Yet there is little use of classroom enactment data to explore this issue more deeply. In this study, the enactments of ICT-supported lessons by two teachers and the performance of their respective classes were analyzed and compared to further unpack the role of the teacher factor in relation to the effectiveness of complex ICT-facilitated lessons.

CSI lessons are intended to create a learner-centred inquiry-based learning environment for students’ individual and collaborative investigation of science phenomena and concepts. In general, the lessons benefited students’ conceptual learning as indicated by the improvement in their test scores. Yet a gap was still identified between the designed and enacted lessons, and this deviation could possibly be explained by teachers’ pedagogical beliefs concerning the instruction. As observed, Charley who possessed more traditional pedagogical beliefs adopted a teacher-guided inquiry approach where more structural and procedural information was provided. Though the general activity pattern enacted was in concert with the designed lesson plan, the way he conducted, assisted, and scaffolded the inquiry and collaborative tasks deviated from what was expected for a good inquiry lesson. This implies that constructivist beliefs may be accompanied by a consistently strict, teacher-directed teaching style if traditional core beliefs of didactic teaching (e.g., regarding discipline and assessment) are still held by the teacher (Teo, 2009). The enactment of student inquiry and collaborative tasks by Katherine was more desirable, yet she failed to deliver a manageable lesson, which led to the different paces of inquiry work conducted by students. This echoes other studies arguing that an effective teacher does not just teach subject content, but also elaborates procedures for completing a task, reaching a goal, solving a problem and making sense of the experience (Rojas-Drummonda & Merce, 2003). The observed design-enactment gap calls for elaborated professional development programs to help teachers master the pedagogical principles of inquiry and CSCL activities and the use of ICTs as cognitive tools (Jonassen, 1995), and to shift their beliefs accordingly.
With the same CSI lesson design, teachers with different pedagogical beliefs and competencies in instructing collaborative inquiry enacted differently. With a traditional pedagogical orientation, the teacher intended to control the class and tended to ignore requests from individual students, yet the teacher with constructivist pedagogical beliefs focused more on scaffolding and elaborating students’ thinking. However, as CSI provides a complex collaborative inquiry approach featuring multiple instructional phrases, teachers are expected to have good understanding of the proposed pedagogical principles, the designed technological affordances, and the specified learning objectives of each inquiry phase to orchestrate the classroom to empower real-time adaptive yet effective enactment of activities leading to desirable processes and outcomes (Chen, Looi, & Chen, 2009). This was neglected in both teachers’ classrooms, with one scaffolding individual and group work insufficiently and the other inadequately attending to students’ skills in collaboration, modeling and reflection. The issues that need addressing include balancing teacher control and the level of student autonomy in inquiry activities (i.e., open inquiry vs guided inquiry); shifting the practices from a focus on logistics to a focus on inquiry (Williams et al., 2004); ensuring that teachers as adaptive facilitators in different inquiry phases; and assisting in students’ inquiry and collaborative activities appropriately.

Implications
Based on above analysis and discussion, implications for ICT-facilitated instruction have been drawn. For teachers whose TEs are more identical to Charley’s (teacher-guided) (Cuban, 1983), we suggest the teachers look for opportunities to intervene more in students’ collaboration and discussion and act as an adaptive facilitator in peer discussion. While moving among groups, teachers should focus more on monitoring students’ understanding than the correctness of answers and/or the completion of tasks. Teachers can interact more with their students to enhance their engagement and to probe their understanding. For teachers whose TEs are more identical to Katherine’s (student-centered), we suggest providing more scripts to them on the purposes and procedures before and during the tasks. The scaffolds should not only be aimed at improving conceptual understanding, but also at the development of regulatory strategies. Teachers need to enculturate students into practices of peer review, collaboration and modeling. Additionally, more exploratory questions need to be asked if most students have difficulty in seeking solutions to problems (Cohen, 1994). For both classes, we suggest that the teachers guide students to conduct more productive and exploratory peer discussions, empowering them to lead and sustain their own dialogue on concepts, prior knowledge and methods. By enabling the students to dialogue together, students will gradually become skilled participants in intellectual communities of discourse and practice (Rummel, Spada, & Hauser, 2009).

Acknowledgments
This research is funded by National Research Foundation in Singapore (NRF2009-IDM001-MOE-019, IDM SST Future School-Science project). We would like to thank the CSI team members and our collaborators: Weikai Fu, Chaohai Chen, Jean Phua Yin Chiu, Shan Gao, Karel Mous, Evelyn Teo, Charles Low, Kassandra Lim, Justin Ke, Kwai Yin Loh and their students for working with us on the project.

References


Exploring the Relationship between Self-Regulated Vocabulary Learning and Web-Based Collaboration

Sarah Hsueh-Jui Liu¹, Yu-Ju Lan²* and Clodia Ya-Yu Ho¹

¹Department of Applied English, Kainan University, Taiwan // ²Department of Applied Chinese Language and Culture, National Taiwan Normal University, Taiwan // sarah.hjliu@gmail.com // yujulan@gmail.com // cloudiaho@hotmail.com

*Corresponding author

(Submitted November 1, 2013; Revised May 13, 2014; Accepted May 24, 2014)

ABSTRACT

Collaborative learning has placed an emphasis on co-constructing knowledge by sharing and negotiating meaning for problem-solving activities, and this cannot be accomplished without governing the self-regulatory processes of students. This study employed a Web-based tool, Google Docs, to determine the effects of Web-based collaboration on vocabulary improvements among learners of English as a foreign language (EFL). The work performed in this study represents a further step toward identifying the factors that influence self-regulated vocabulary strategy use and perceptions of Web-based collaboration (SRvsWBC). In total, 210 undergraduate students participated in this study and undertook the designed tasks, such as vocabulary pre-/posttests and a self-report questionnaire survey of SRvsWBC. The findings of the study suggest that collaboration using a Web-based tool affects knowledge development, and provide insights into the integrated spectrum of self-regulation, L2/FL learning, and Web-based technology that will be useful for pedagogy.

Keywords

Web-based collaboration, English as a foreign language (EFL), Vocabulary strategy use, Self-regulated learning (SRL)

Introduction

Accessing the Internet has become the most popular means by which college students communicate with one another on a daily basis, such as via email (Lee, Cheung, & Chen, 2005). Such Web 2.0 social tools have been considered to support the meaningful and collaborative learning (Gao, 2013; Tsai & Tsai, 2013) that underlies a social cognitive view of self-regulated learning (SRL). These tools reportedly motivate students to become independent learners who are able to direct and take control of their own learning and to evaluate their learning goals in order to acquire knowledge and skills (Pintrich, Smith, Garcia, & McKeachie, 1993; Zimmerman, 1990). There has been a substantial amount of research into SRL in an attempt to examine the importance of an individual’s internal (e.g., interests and motivation) and external (e.g., interaction between interlocutors) characteristics in both face-to-face settings and technology-based environments (e.g., Cho & Jonassen, 2009; Cho & Kim, 2013; Schmidt, Boraie, & Kassabgy, 1996). It has been stated that those who are seeking language-use opportunities outside the classroom in technology-based settings “have been found to be more likely to take responsibilities for self-directing and self-managing their language learning and use technologies to regulate their language learning experience” (Lai, 2013, p. 103). Understanding the internal constructs in relation to contextual factors (e.g., learning mathematics) is essential, but there has been little research in this area (Hsin, Lin, & Yeh, 2005).

Building on previous work, this study aimed to provide a better understanding of whether the use of Web-based collaboration tool influences the vocabulary improvements of learners and whether the subscales of underlying variables are identifiable and interrelated. This study neither analyzed the negotiated interaction of the participants where they were using a Web-based tool (Google Docs in this case) nor assessed the effectiveness of the learning activities undertaken by the participants; instead, the aim was to measure the impact of collaboration on the vocabulary improvements of the participants in a Web-based environment. This study further aimed to identify the relationship between the constructs under investigation.
Literature review

Self-regulated second-language or vocabulary learning

SRL refers to individuals who “are metacognitively, motivationally, and behaviorally active participants in their own learning process” (Zimmerman, 1989, p. 329). This posits that SRL involves “not only cognitive but also motivational and affective factors, as well as social contextual factors” (Pintrich, 2004, p. 386). Broadening the scope of SRL in educational psychology to encompass second or foreign language (L2/FL) learning, learners can be said to actively employ strategic tactics to control their emotions (e.g., being anxious when conversing with native speakers of the L2) and to overcome unknown vocabulary or linguistic deficits in order to facilitate effective communication (Oxford, 2011).

Affective constructs, including motivation and attitudes toward L2 learning, have been empirically examined in relation to the use of L2 learning strategies (e.g., Schmidt, Boraie, & Kassabgy, 1996). It is believed that students with higher levels of motivation display higher quality learning strategies and outcomes. This is consistent with language strategy use being found to be inextricably intertwined with the motivation or self-perceived ability to determine the degree of outcomes. L2 learners with greater motivation are likely to employ more types of strategies, such as communication-oriented strategies, and to do this more frequently (Oxford & Nyikos, 1989). Instrumentally motivated students who were inclined to learn an L2 for practical reasons were found to prefer using strategies in connection with a first language and other languages more frequently, whereas integrative motivated students tended to learn the language just out of interest (Levine, Reves, & Leaver, 1996).

In the context of L2 vocabulary learning, because word knowledge plays an important role in receptive and productive skills in association with effective communication (Liu, Lan, & Jenkins, in press; Nyikos & Fan, 2007), insufficient vocabulary knowledge will cause learners to experience communication breakdowns when confronted with unknown words or expressions. Tseng, Dörnyei, and Schmitt (2006) proposed an integrative model of SRL and vocabulary strategy use based on the conceptual framework of SRL. Instead of focusing at a micro level of specific vocabulary strategy use, such as paying attention to the prefix and suffix of a word, their model emphasizes that a learner’s innate self-regulatory capacity enables him- or herself to choose appropriate strategies to activate a macro level of learning. Such general aspects involve an individual’s commitment control (to achieve learning goals), metacognition control (to manage concentration on learning), satiation control (to eliminate boredom), and emotional and environmental control. The relations between vocabulary strategy use and motivation are also evident from vocabulary interventional studies. Mizumoto and Takeuchi (2009) found that the motivation of students could be enhanced by explicitly teaching them vocabulary strategies inside the classroom, such as reviewing learned vocabulary (metacognition strategies) and making a mental picture (cognitive strategies).

SRL and web-based technology for language learning

Web-based technology can support SRL in L2 learning, since it “can shift control to the learner, through promoting learner agency, autonomy, and engagement” (McLoughlin & Lee, 2010, p. 28). Such Web-based tools for L2 learning contexts underlie the sociocultural theory (cf. Lantolf & Thorne, 2007; Vygotsky, 1978) that assumes that cognitive developmental processes take place through collaborative interaction between interlocutors in a social milieu, whereby learners are engaged in co-constructing their knowledge in the target language (Swain & Lapkin, 2000). Various research studies have investigated the multifaceted dimensions of how use of technology improves L2 learning. While some studies have examined meaningful interactions between experienced and novice interlocutors (e.g., Sotillo, 2005; Yılmaz & Granena, 2010), others have focused on how such technology allows students to work collaboratively in order to enhance learning experiences (e.g., Chang, 2005; Gao, 2013; Kessler, Bikowski, & Boggs, 2012; Lai & Gu, 2011).

Cheng and Chau (2013) examined the relationship between the SRL ability of students and their e-portfolio achievements, and found that use of cognitive and self-regulated strategies was positively related to learning outcomes. Their scoring of e-portfolios revealed that higher achievers employed more cognitive processing strategies, such as elaboration, organization, and critical thinking, and metacognitive self-regulation and collaborative peer-learning strategies. Chang (2005) examined the relations between SRL strategies and how learners perceive their
motivation for Web-based instruction. The results of that study indicate that learners can improve their learning motivation and confidence by incorporating SRL strategies and Web-based instruction.

Regarding the use of a Web-based medium for the out-of-classroom support of student learning, Lee et al. (2005) attempted to measure students’ acceptance of such a learning medium within a motivational framework in terms of two main factors: extrinsic (perceived usefulness and ease of use) and intrinsic (perceived enjoyment). The findings show that apart from perceived ease of use, both extrinsic motivation and intrinsic motivation are significantly related to the attitudes of students and their intention to use a Web-based learning medium. As hypothesized, motivational subscales, such as extrinsic and intrinsic, are considered to be associated with how students perceive the use of technology.

**Use of web-based technology for language learning**

Web-based technology provides opportunities for authentic learning and meaningful communication (Kabilan, Ahmad, & Abidin, 2010). For example, Warschauer (1996) investigated the difference in participation by learners between electronic and face-to-face discussion in L2/FL learning. The results of his study reveal that learners exhibited more equal opportunities when expressing their thoughts in the electronic modality than in the face-to-face modality. Learners not only produced more complex lexical and syntactic structures, but also had a more positive attitude toward discussing issues electronically. Kessler et al. (2012) assessed collaborative writing by learners that focused on linguistic features (e.g., verb tenses and meaning), strategic processing (e.g., planning), and the perception of the collaboration when using the Google Docs Web-based tool. Their results suggest that this Web-based tool allowed students to exhibit their abilities in using collaborative scaffolding in relation to the forms and meanings of words, employing multiple strategies in writing, such as editing and revising, as well as feeling positive toward working in such an environment.

One of important factors influencing language learning is how learners perceive Web-based collaboration that may improve their learning in the target language. Davis (1989) suggested that perceived usefulness and ease of use are the two factors underlying the degree to which users believe that using technology will enhance their performance. Both of these factors relate to an individual’s cognitive responses to using technology, which will in turn influence the user’s emotional states (Holden & Rada, 2011). Such feelings and opinions regarding the perception of technology use outside the classroom have been investigated empirically, with the results showing that learners generally have positive attitudes toward technology use and acceptance (Lai, Wang, & Lei, 2012; Lee et al. 2005). As hypothesized in the present study, the L2 vocabulary strategy use by learners to enhance English learning is associated with their acceptance and perception of Web-based collaboration.

The participants in this study worked together as group members on Google Docs either at the same time or at different times. However, since collaborative work involves individuals working in a group (Stahl, Koschmann, & Suthers, 2006), completing problem-solving tasks or working collaboratively can be problematic (Chisholm, 1990). For example, some individuals in a particular group might not feel comfortable giving their opinions or commenting on the work of others due to a lack of confidence in their English ability, whereas other individuals may be more willing to work with others for various reasons, such as in order to complete the tasks to help them get a good grade. How individuals perceive collaborative learning is reflected in their behaviors, and these were used in the present study to identify them as active or passive collaborators. Together with the theoretical concepts relevant to SRL and technology use, this study attempted to answer the following questions: (1) is there a significant relationship between the scores in pre- and posttests, (2) is there a significant difference in the vocabulary improvements between active and passive collaborators, and (3) what factors influence self-regulated vocabulary strategy use and the perception of Web-based collaboration (SRvsWBC)?

**Method**

**Participants**

In total, 210 first- and second-year university students from 6 intact English classes at a private university in northern Taiwan were recruited for this research. The final number of included participants was 180 due to 16 students being
absent during the administration of the questionnaire survey and 14 students completing less than half of the questionnaire. The number of female participants \(N = 114\) was nearly double that of male participants \(N = 66\), please see Table 1. The students voluntarily participated in this study, and their age ranged between 19 and 26 years old. They were studying in various fields and shared similar backgrounds, speaking Chinese or Taiwanese as a first language and learning English as a core subject and as an FL ever since when they had been elementary-school students.

The demographic information of the participants shown in Table 1 indicates that almost all of the students (97%) were aged between 19 and 21 years, with only 3% being aged between 22 and 26 years. Furthermore, more than half of the students were in their first year of studies, whilst approximately 40% of them were second-year students. Majors in Tourism and Hospitality comprised the largest proportion of students (37%), whereas the majors in Logistics and Shipping Management comprised the smallest proportion (11%).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>66</th>
<th>(37%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>114</td>
<td>(63%)</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>(100%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>19-21 years</th>
<th>138</th>
<th>(97%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22-26 years</td>
<td>5</td>
<td>(3%)</td>
</tr>
<tr>
<td>Total</td>
<td>143</td>
<td></td>
<td>(100%)</td>
</tr>
<tr>
<td>37 missing cases</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Year of study | First year | 102 | (57%) |
|               | Second year | 78  | (43%) |
|               | Total       | 180 | (100%)|

| Major subject | Tourism and Transportation | 23  | (13%) |
|               | Logistics and Shipping Management | 19  | (11%) |
|               | Tourism and Hospitality | 67  | (37%) |
|               | Leisure and Recreation Management | 25  | (14%) |
|               | Air Transportation | 46  | (25%) |
|               | Total | 180  | (100%)|

| Length of time spent learning English | 7-8 years | 78  | (55%) |
|                                       | 9+ years  | 64  | (45%) |
|                                       | Total     | 142 | (100%)|
| 38 missing cases |           | |

| Length of time spent in English-speaking countries | None | 148 | (92%) |
|                                                     | 1–6 months | 11  | (7%)  |
|                                                     | < 1 year    | 2   | (1%)  |
|                                                     | Total       | 161 | (100%)|
| 19 missing cases |           | |

Regarding the length of time spent learning English, the homogeneous group of students started to learn English as a school core subject when they were in sixth grade at elementary school. This meant that approximately half of the students had been learning English for more than 7 years. A mere 1% of students had visited an English-speaking country, such as the United Kingdom or the United States.

**Instruments and materials**

**Vocabulary test**

The English vocabulary test used to measure the participants’ vocabulary knowledge comprised a total of 40 items divided into 5 sections. Sections A and C, respectively comprising five and eight questions, involved matching tests between English words with similar meanings, such as “exercise” being akin to “work out.” Section B, comprising five questions, provided five target lexical items with five gaps in a short reading passage. Similar to Section C, Section D (12 questions) involved matching English words with their meaning in Chinese. In the last section, ten target items underlined in a text were tested in ten multiple-choice questions. The designed content of the pre- and posttests was mainly based on the participants’ textbook.
Self-report questionnaire survey

The paper-and-pencil-administered questionnaire survey of SRvsWBC, translated from English into Chinese, consisted of two sections. The first section contained question items designed to obtain learners’ demographic information, such as their majors, gender, and age. The second section comprised 3 parts with a total of 52 items scored on a 5-point Likert scale, ranging from “strongly disagree” (score of 1) to “strongly agree” (score of 5). Part 1 measured the motivational beliefs of learners based on Zimmerman (1989), Pintrich and de Groot (1990), and Gardner (1985); the total of 20 items were divided into the following 3 subscales: (1) self-efficacy (6 items), dealing with one’s perception and beliefs for learning, such as “compared with others, I think I’m a good student”; (2) motivation (10 items), focusing on reasons for engagement in academic tasks, such as “I like doing English tasks”; and (3) test anxiety (four items), such as “I have an uneasy, upset feeling before taking a test.”

Part 2 assessed how learners employ vocabulary strategies in L2 using 14 items designed based on the taxonomies of Schmitt (2000) and Oxford (1990), and involved the following 4 factors: (1) determination (3 items), referring to discovering the meaning of a new word, such as “paying attention to the structures of new words”; (2) social strategies (3 items), relating to tactics involving interacting with others, such as asking for help from a teacher or peers; (3) memory (5 items), regarding mnemonics methods using grouping or imagery, such as “listing words that are related to each other”; and (4) metacognition strategies (3 items), involving a series of conscious actions including planning, monitoring, and evaluating an individual’s learning, such as “reviewing new words.”

Part 3 comprised 18 items on the perception of Web-based collaboration based on the work of Davis (1989), and was divided into 2 subscales: (1) perceived usefulness, comprising 11 items such as “using the Web-based tool increases my English vocabulary ability”; and (2) ease of use, comprising 7 items such as “I find Google Docs easy to use.” Each subscale consisted of four questions.

Learning materials

Five reading texts taken from the participants’ textbooks and online reading materials (e.g., English Club) for use as vocabulary learning materials were designed for the participants to practice and work collaboratively on Google Docs. Each text contained fewer than 500 words and included some questions to be answered after the text had been read. For example, in one of the five articles regarding the impact of fast food on people’s health, the questions related to reading comprehension included “what impression do people have about fast food according to this article” and “please list a number of new vocabulary words that appear in this article and how you will remember them.”

Figures 1 and 2 provide examples of synchronicity and asynchronicity to illustrate how the collaborators worked on Google Docs. Figure 1 shows an example of a learning activity performed asynchronously by a group of students on Google Docs, whereas Figure 2 provides an example of two collaborators working synchronously. Changes or attempted changes marked in different colors, which indicates activities performed by different members, are recorded in a history list in the right-hand column, with the colors allocated to the students selected randomly by the system. For example, Figure 1 shows that clicking the student in blue makes her answer to the comprehension question appear (in blue) in the central area.

Figure 1. An example of group members collaborating asynchronously on a learning activity
Google docs

The current project provided participants with access to the learning materials via Google Docs, which were designed to supplement the compulsory English class in which the students and the researcher met for 2 hours a week in a face-to-face classroom setting. Google Docs provides many free tools similar to those included in the Microsoft Office software package (e.g., Word, Excel, and PowerPoint) that allow multiple users to create/edit documents or share files and collaborate with others. In this research the participants collaborated synchronously or asynchronously on given tasks using the Word processing tool.

Procedure

There were two workshop sessions conducted for the participating classes at a computer laboratory in the university (see Figure 3). The first session, held in the first week of the semester, was run as a tutorial. Each class of participants was divided into small groups, each comprising four or five students, and a new Gmail account was created for each of them to allow them to access Google Docs. The purpose of the first workshop was essentially to help the participating students to familiarize themselves with the tool functions on Google Docs, such as uploading a file, creating and sharing a document, and collaborating with others. The second session, conducted in the middle of the semester, served two purposes: (1) to check whether the collaborators had encountered any technical obstacles when using the Web-based tool and (2) to check whether the collaborators were having any problems using the tools provided on Google Docs. Each session lasted 50 minutes.

In the first workshop session the students also completed a vocabulary pretest. Following this, small groups of students started practicing vocabulary learning by reading texts on Google Docs when they were at home or on campus. They were allowed to provide their answers to the questions and to comment on the answers provided by others, and they were reminded to use English as much as possible when answering the questions. They were encouraged to collaborate with their group members when answering the post-reading questions (as discussed in the section above) by using the vocabulary strategies taught in the face-to-face setting of the classroom. Note that the participants were taught vocabulary strategies in that face-to-face setting, including mnemonic or memory methods. The types of strategies taught in class were entirely dependent on the contents of their textbook. Five reading texts in total were used for the vocabulary learning on Google Docs. A new text was uploaded approximately every 3 weeks. This protocol resulted in three articles being available in the system before the second workshop took place in week 9, so that students could provide feedback about using the system. In the last week the students took the vocabulary posttest and completed the SRvsWBC questionnaire.
Data analysis

A paired-samples t-test using SPSS version 17 was performed to examine the effect of the Web-based collaboration tool on the vocabulary improvements. In addition, two types of collaboration were identified according to the individual contributions to a given task: passive and active collaboration. The learning activity entitled “Fast Food” was randomly selected for the analysis of these two types of collaboration. The criteria of this classification were based on the collaborative effort made by the participants, such as expressing personal prior experience (as see the details in Table 3). All attempts made by the participants were coded independently by the two researchers of this study; the results showed that the interrater agreement was 86% in the data analysis.

Factor analysis and Pearson’s product-moment correlation were computed to determine the subscales of the latent constructs, e.g., vocabulary strategy use, and to identify the relationships between the underlying constructs. The purpose of the correlation analysis was to determine whether the variables being investigated were positively related, negatively related, or not related.

Results

English vocabulary improvements

A paired-samples t-test was conducted to measure the impact of the Web-based collaboration tool on learners’ vocabulary scores between the pre- and posttests (see Table 2). The mean score in the English vocabulary test increased significantly from the pretest (Mean = 23.22, SD = 7.44) to the posttest (Mean = 29.41, SD = 7.55, t(137)
The increased scores in the posttest suggest that the participants improved their vocabulary knowledge by learning collaboratively in the Web-based environment over the 18-week study period. These results indicate a significant relationship between the vocabulary scores in the pre- and posttests.

We also examined how the participants collaborated with others within their own group. Active collaborators were those who contributed by providing their answers to questions and discussing questions with others or by commenting on the answers provided by others (see Table 3).

### Table 3. Criteria of active collaborators

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Examples of students’ statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Expressing personal prior experience</td>
<td>“I think the fast food is very convenience to buy it but don’t eat too much. They are not good for your health, because you don’t know what make it look delicious. If you can, you should cook a meal by yourself. It will better for your health.” (S29)</td>
</tr>
<tr>
<td>2. Being engaging in discussions</td>
<td>“Yes, I think so.” (S27)</td>
</tr>
<tr>
<td>3. Commenting on others’ answers</td>
<td>“I don’t think there is any advantages of fast food, the only advantage is very convenient to buy whenever you feel hungry. When the first time I saw the processes of making fast food, I told to myself never ate fast food any more. High fat and high sugar will lead us to death.” (S24)</td>
</tr>
<tr>
<td>4. Providing personal opinions or suggestions</td>
<td></td>
</tr>
</tbody>
</table>

Note. “S” indicates “Student no.”

One of questions asked of the students for the “Fast Food” topic was “What are the advantages or disadvantages of fast food?” Although the statement made by Student no. 29 (S29) contained grammatically incorrect sentences and inappropriate expressions, such as “…is very convenience to buy…,” the statement provided his personal opinions and suggestions, such as “…because you don’t know what make it look delicious” indicating that the ingredients of fast food might not be healthy to eat. He suggested that people should prepare meals themselves so as to ensure that they eat healthy food. S27 commented on what S29 had stated by saying, “yes, I think so too,” though it was not clearly indicated which part of the statement made by S29 she was agreeing with. However, S24 provided a more complete answer than that of S29, in that while S29 said that fast food was convenient to buy, S24 explained why this was the case: that it was convenient to buy when one is hungry. S24 also expressed her personal experience and emotions when she saw how people make the fast food, and this made her determined not to eat fast food again.

In contrast to active collaboration, passive collaboration refers to not providing any answers to the questions as well as only viewing the work of others, without making any contributions to the collaborative work. For example, S28 and S25 (who were in the same group) only viewed the responses without providing any answers.

According to the two types of collaboration discussed above, a further statistical analysis was conducted with the aim of identifying differences in the vocabulary pre- and posttests between passive (Group A) and active (Group B) collaborators. This revealed that there were more passive collaborators (N = 84) than active collaborators (N = 64 or 60). Once again, this data analysis was based on one of five randomly selected learning tasks.

### Table 4. Results of paired-samples t-tests for pre- and posttests between groups

<table>
<thead>
<tr>
<th></th>
<th>Group A</th>
<th>Group B</th>
<th>t</th>
<th>p (two-tailed)</th>
<th>eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>84</td>
<td>64</td>
<td>7.2</td>
<td>.003</td>
<td>.04</td>
</tr>
<tr>
<td>Posttest</td>
<td>84</td>
<td>60</td>
<td>2.4</td>
<td>.01</td>
<td>.04</td>
</tr>
</tbody>
</table>
Table 4 lists the results of a paired-samples t-test conducted to measure the impact of collaborative learning on the vocabulary improvements of the participants. The table indicates that there was no significant group difference ($p > .05$) in the pretest, whereas there was a significant difference in the posttest between Group A (Mean = 22.52, $SD = 7.58$) and Group B [Mean = 27.65, $SD = 7.79$, $t(142) = 2.43$, $p < .01$]. The eta-squared value (.04) indicates the presence of a small-to-moderate effect size.

The above results suggest that there was a significant difference in vocabulary improvements between the pre- and posttests ($p < .000$). Moreover, there was a significant association in the posttest between the passive and active collaborators ($p < .01$).

**Factor analysis for identifying the underlying constructs**

To identify the constructs of SRvsWBC, an exploratory factor analysis was performed to explore the interrelationship among a set of subscales in each construct. The maximum likelihood and promax were employed when using the extraction and rotation methods. In accordance with Pallant (2005) and Field (2009), the criteria for determining the suitability of extracting factors were (1) correlation coefficients of at least .4, (2) Kaiser-Meyer-Olkin (KMO) values exceeding .5, (3) the scree plot indicating the factors to be extracted, (4) Barlett’s Test of Sphericity reaching a significance level of .000, and (5) eigenvalues exceeding Kaiser’s criterion of 1.

In the construct of motivational beliefs, the results suggest four subscales or factors, including intrinsic and extrinsic motivations, test anxiety, and self-efficacy (see Table 5). However, the two items on the subscale of self-efficacy (Q4, “I believe that I can keep up with what the teacher teaches”; and Q18, “compared with others, I think I have good study skills”) were eliminated since they did not contribute to any factors. Table 5 indicates that the factor loadings ranged between .52 and .97. The results suggest that the items clustered in Factor 1 represented intrinsic motivation, with eight items involving goals and beliefs that any given learning activity was interesting and essential, such as “I like to take challenging English tasks.” Factor 2 represented test anxiety, comprising four items such as “I tend to think about how poorly I am doing.” Factors 3 and 4 represented self-efficacy (four items) and extrinsic motivation (two items), respectively. These four factors explained 62.75% of the total variance in this construct. The value for the KMO Measure of Sampling Adequacy was .827, Bartlett’s Test of Sphericity produced $x^2(171) = 1314.08$ and $p < .000$, and Cronbach’s $\alpha$ coefficients exceeded .70, which indicates sound reliability.

All responses in the motivational beliefs were rated on a 5-point Likert scale, ranging from “strongly disagree” (score of 1) to “strongly agree” (score of 5). Table 5 indicates that the participants made positive statements regarding intrinsic motivation, self-efficacy, and extrinsic motivation; in particular, self-efficacy had the highest mean score with 4.23, which indicated that those who had higher self-efficacy for learning were more likely to think they would get a better grade or to think they knew more about the subject than others did. In contrast, participants responded less positively regarding test anxiety; the mean score was 2.97, which suggested that the participants tended to agree that they experienced only mild test anxiety.

**Table 5. Summary of results of exploratory factor analysis for the questionnaire survey of motivational beliefs ($N = 146$)**

<table>
<thead>
<tr>
<th>Item statement</th>
<th>INT</th>
<th>TES</th>
<th>SEL</th>
<th>EXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9. I like to take challenging English tasks</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11. I like brainstorming for English practice tasks</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q20. I complete English tasks by myself</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12. I like explaining to others about what I know</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q10. I like doing English tasks</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3. I am confident in English learning</td>
<td>.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q19. I accomplish tasks without looking at the answers first</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q17. I like learning new things from a learning task</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13. I tend to think about how poorly I am doing</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14. I worry about tests before taking one</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15. I think about the questions that I cannot answer</td>
<td>.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16. I have an uneasy, upset feeling before taking a test</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q5. Compared with others, I think I'm a good student  .91
Q6. I can do well in the classroom  .84
Q1. I know a great deal about this subject  .58
Q2. I'll get a good grade  .58
Q8. Better English skills will help me to get a better job  .97
Q7. People will respect me if I have an excellent English ability  .61

Eigenvalue  
6.00  2.92  1.72  1.29
Percentage of total variance  
31.58  15.35  9.04  6.78
Cumulative variance (%)  
62.75
Cronbach’s α  
.86  .86  .82  .71
Mean  
3.30  2.97  4.23  3.50
SD  
4.78  3.25  2.43  1.41

Note. Abbreviations: INT = intrinsic motivation, TES = test anxiety, SEL = self-efficacy, EXT = extrinsic motivation.

Table 6. Summary of results of exploratory factor analysis for vocabulary learning strategies (N = 140)

<table>
<thead>
<tr>
<th>Item statement</th>
<th>Factor 1: MEM</th>
<th>Factor 2: MET</th>
<th>Factor 3: DET</th>
<th>Factor 4: SOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q14. Connecting words to one’s experience</td>
<td>.86</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6. Using semantic maps to help remember new words</td>
<td>.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12. Using keyword methods to remember words</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7. Listing words that are related to each other</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13. Studying the spelling of a word</td>
<td>.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9. Reviewing new words</td>
<td></td>
<td></td>
<td></td>
<td>.94</td>
</tr>
<tr>
<td>Q11. Previewing what is to be learned</td>
<td>.68</td>
<td></td>
<td></td>
<td>.60</td>
</tr>
<tr>
<td>Q10. Testing oneself with new words</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1. Paying attention to the structures of new words</td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3. Grouping words according to the parts of speech</td>
<td></td>
<td>.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2. Breaking a word into small parts that I can remember</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5. Asking for help from a teacher</td>
<td></td>
<td></td>
<td></td>
<td>.94</td>
</tr>
<tr>
<td>Q4. Asking for help from peers</td>
<td></td>
<td></td>
<td></td>
<td>.60</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>5.83</td>
<td>1.65</td>
<td>1.30</td>
<td>1.11</td>
</tr>
<tr>
<td>Percentage of total variance</td>
<td>36.42</td>
<td>10.30</td>
<td>8.13</td>
<td>6.97</td>
</tr>
<tr>
<td>Cumulative variance (%)</td>
<td></td>
<td></td>
<td></td>
<td>61.81</td>
</tr>
<tr>
<td>Cronbach’s α</td>
<td>.82</td>
<td>.80</td>
<td>.74</td>
<td>.68</td>
</tr>
<tr>
<td>Mean</td>
<td>3.21</td>
<td>3.55</td>
<td>3.50</td>
<td>2.99</td>
</tr>
<tr>
<td>SD</td>
<td>3.19</td>
<td>1.74</td>
<td>1.87</td>
<td>1.57</td>
</tr>
</tbody>
</table>

Note. Abbreviations: MEM = memory, MET = metacognition, DET = determination, SOC = social.

Table 6 indicates that the analysis of the vocabulary strategy use yielded the following four factors: (1) memory, comprising five items such as “connecting words to one’s experience” and “using semantic maps to help remember new words”; (2) metacognition, comprising three items such as “reviewing words that are related to each other,” “previewing what is to be learned,” and “testing oneself with new words”; (3) determination, comprising the three items of “paying attention to the structures of new words,” “grouping words according to the parts of speech,” and “breaking a word into small parts that I can remember”; and (4) social, comprising the two items of “asking for help from a teacher” and “asking for help from peers.” Factors 1, 2, 3, and 4 explained 36.42%, 10.30%, 8.13%, and 6.97% of the total variance, respectively. The value of the KMO Measure of Sampling Adequacy was .85, and Bartlett’s Test of Sphericity produced χ²(120) = 865.79 and p < .000. All values of Cronbach’s α apart from that for Factor 4 (social strategy) exceeded the recommended value of .70.

One item, a social factor (Q8, “Study and practice meaning in a group”), was discarded because its coefficient value was less than .40. Table 6 indicates that the four factor loadings ranged between .44 and .94. In terms of the vocabulary strategies that were most frequently employed, the metacognition and determination factors had similar mean scores of 3.55 and 3.50, which means that the participants preferred using metacognition and determination strategies when learning vocabulary words. However, the social factor had the lowest mean of 2.99, which means the students tended not to ask for help from their teachers or peers when they were experiencing difficulty in learning lexical words in English.
Table 7 presents the results for the perception of the Web-based collaboration survey. This survey identified two factors: (1) usefulness, comprising nine items such as “Google Docs helps me accomplish my English tasks”; and (2) ease of use, comprising six items such as “it is easy to remember how to use Google Docs.” Factors 1 and 2 explained 59.46% and 7.50% of the total variance, respectively. The value of the KMO Measure of Sampling Adequacy was .95, Bartlett’s Test of Sphericity produced X²(120) = 1763.41 and p < .000, and all values of Cronbach’s α exceeded .90, indicating good internal consistency.

### Table 7. Summary of results of exploratory factor analysis (N = 145)

<table>
<thead>
<tr>
<th>Item statement</th>
<th>Factor 1: USE</th>
<th>Factor 2: EAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q7. Google Docs helps me accomplish my English tasks</td>
<td>.85</td>
<td></td>
</tr>
<tr>
<td>Q17. Learning activities improve my English ability</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td>Q10. Using the Web-based tool increases my English vocabulary ability</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Q4. Learning activities are helpful for learning vocabulary words</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>Q18. Learning tasks are helpful for increasing vocabulary knowledge</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Q6. Learning tasks increase my interest in learning English</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Q16. Learning tasks increase my reading ability</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Q11. Google Docs is an effective learning environment</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>Q1. Google Docs is an effective tool for learning English</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Q3. Google Docs is easy to use</td>
<td></td>
<td>.95</td>
</tr>
<tr>
<td>Q9. I became familiar with the system by learning how to use it</td>
<td></td>
<td>.90</td>
</tr>
<tr>
<td>Q15. It is easy to remember how to use Google Docs</td>
<td></td>
<td>.67</td>
</tr>
<tr>
<td>Q2. I like to use this system for learning English</td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>Q13. I find it easy to use the system after I was first taught how to</td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>Q8. When using Google Docs, I don’t consult anyone else</td>
<td></td>
<td>.43</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>9.51</td>
<td>1.20</td>
</tr>
<tr>
<td>Percentage of total variance</td>
<td>59.46</td>
<td>7.50</td>
</tr>
<tr>
<td>Cumulative variance (%)</td>
<td></td>
<td>66.96</td>
</tr>
<tr>
<td>Cronbach’s α</td>
<td>.95</td>
<td>.90</td>
</tr>
<tr>
<td>Mean</td>
<td>3.55</td>
<td>4.00</td>
</tr>
<tr>
<td>SD</td>
<td>5.62</td>
<td>4.41</td>
</tr>
</tbody>
</table>

Note. Abbreviations: USE = usefulness, EAS = ease of use.

Two items in the usefulness factor (Q5, “Without the system, it is difficult to perform my job”; and Q12, “I will use the system for other subjects”) and one item in the ease-of-use factor (Q14, “I often make errors when using the system”) were eliminated for the following reasons: Q5 and Q12 did not cluster into a single factor, and Q14 had a coefficient less than .4. The first-factor loading had a smaller coefficient range of between .70 and .85, whilst the second-factor loading had a wider coefficient range of between .43 and .95. The mean scores for the usefulness and ease-of-use factors were 3.55 and 4.00, respectively, which indicates that the students had a positive attitude toward using the Web-based tool.

The above results show that three constructs are associated with SRvsWBC: (1) motivational beliefs, comprising intrinsic motivation, extrinsic motivation, self-belief, and test anxiety; (2) vocabulary strategy use, comprising memory, metacognition, determination, and social strategies; and (3) perception of Web-based collaboration, comprising usefulness and ease of use. These constructs as factors that influence SRvsWBC were co-related, as described below.

### Interrelations of the constructs

The interrelationships among the subscales of motivational beliefs, vocabulary strategy use, and perception of Web-based collaboration were examined using Pearson’s product-moment correlation. A preliminary analysis indicated that there was no violation of the assumptions of normality, linearity, and homoscedasticity (Pallant, 2005). First, as indicated in Table 8, there was a strong positive correlation between the two factors of usefulness and ease of use (r = .78, N = 145, p < .01). The second-highest coefficients loaded between intrinsic motivation and self-efficacy (r = .51), between memory strategy and determination strategy (r = .55), and between intrinsic motivation and
memory strategy \(r = .56\). In addition, there were two sets of lowest coefficients (both \(r = .18\)), showing significant associations between ease of use and self-efficacy and between memory strategy and usefulness.

Second, Factor 1 (intrinsic motivation) was significantly associated with all factors except for Factors 2 (test anxiety) and 8 (social), with coefficients ranging between .26 and .56. The results imply that students who had higher intrinsic motivation were more likely to employ vocabulary strategies (e.g., memory, metacognition, and determination strategies). Despite the tendency for students with greater intrinsic motivation to employ a greater number and a wider variety of vocabulary strategies, they were unlikely to use social tactics, such as interacting with their peers or their teacher when learning vocabulary. Furthermore, the intrinsic type of students had a positive attitude toward learning on Google Docs.

Third, Factor 3 (self-efficacy) was significantly related to all factors except for Factor 8 (social), implying that higher self-efficacy was significantly associated with greater use of memory, metacognition, and determination strategies, and a better perception of Web-based collaboration. In addition, Factor 5 (memory) was significantly associated with all factors except for Factor 10 (ease of use), suggesting that those who had motivational beliefs tended to employ four types of vocabulary strategies and perceived that using Google Docs could enhance their learning of English vocabulary. Lastly, Factor 2 (test anxiety) was not significantly related to any factors; similar to Factor 8 (social), it was not significantly associated with any factors except with Factor 5 (memory), as indicated by the correlation coefficient between the two factors being .30.

Table 8. Pearson’s correlations between the subscales

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INT</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. TES</td>
<td>.08</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. SEL</td>
<td>.51**</td>
<td>.14</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. EXT</td>
<td>.44**</td>
<td>.06</td>
<td>.29**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. MEM</td>
<td>.56**</td>
<td>.04</td>
<td>.41**</td>
<td>.24**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. MET</td>
<td>.42**</td>
<td>.05</td>
<td>.38**</td>
<td>.15</td>
<td>.50**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. DET</td>
<td>.38**</td>
<td>-.01</td>
<td>.31**</td>
<td>.24**</td>
<td>.55**</td>
<td>.47**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. SOC</td>
<td>.05</td>
<td>.16</td>
<td>.11</td>
<td>.09</td>
<td>.30**</td>
<td>.14</td>
<td>.16</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. USE</td>
<td>.29**</td>
<td>.10</td>
<td>.20**</td>
<td>.23**</td>
<td>.18**</td>
<td>.10</td>
<td>.09</td>
<td>.05</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>10. EAS</td>
<td>.26**</td>
<td>.13</td>
<td>.18**</td>
<td>.29**</td>
<td>.14</td>
<td>.03</td>
<td>.06</td>
<td>.01</td>
<td>.78**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes. \(N = 146\), \(p < .05\). (two-tailed), \(*p < .01\). (two-tailed).  

Discussion and conclusions

The aim of this study was to determine how Web-based collaboration influences vocabulary improvements in students of English as an FL and to elucidate the relationships among the subscales of SRvsWBC. A number of statistical results are provided. The first shows that the mean scores in the posttest differed significantly from those in the pretest. Moreover, two types of collaboration that were identified by examining participants’ efforts and their types of contribution (passive and active) in one learning task were further analyzed statistically. The results suggest that active collaborators obtained better vocabulary knowledge than did passive collaborators. Such results corroborate the finding of Mendenhall and Johnson (2010) that the use of an online social annotation tool can improve the academic literacy of students who collaborate in small groups. Similarly, Hwang et al. (2007) found that using Web-based annotation tools was associated with learning achievement, in that students who used such tools outperformed those who did not.

The results of our exploratory factor analysis for identifying the subscales of SRvsWBC indicated the good internal consistency of each subscale, with Cronbach’s \(\alpha\) values ranging between .71 and .95 (with the exception of the social factor in vocabulary strategy use). As predicted, the three theoretical constructs each contained several subscales; (1) motivational beliefs comprised positive motivational orientations, self-perceived ability, and negative test anxiety; (2) vocabulary strategy use comprised memory, metacognition, determination, and social strategies; and (3) perceptions included usefulness and ease of use. It is difficult to compare between the constructs used in different studies since they have examined different constructs; however, to some extent the findings of the current investigation; specifically, that an individual’s cognition, motivation, and behavior are interrelated which are supported by the
findings of previous studies, including those of Tseng et al. (2006), Schmidt et al. (1996), and Cho and Jonassen (2009).

It is worth noting that, as previous research has suggested, motivational beliefs involve self-efficacy, intrinsic value, and test anxiety (Pintrich & de Groot, 1990). However, the factorability of the 18 items suggested that they could be subdivided into 4 distinct factors: intrinsic motivation, extrinsic motivation, test anxiety, and self-efficacy. In particular, splitting the motivation item into two subscales was supported by theoretical assumptions in L2 acquisition, such as has been reported by Gardner and Lambert (1972) and Dörnyei (1990), and whereas intrinsic motivation as an integrative-oriented motive refers to learners’ desire for and interest in learning the target language, the extrinsic motivation acted as an instrumental motive, whereby the students studied the target language for some utilitarian purposes, such as obtaining a better job.

The positive internal aspect of motivational beliefs (intrinsic motivation, extrinsic motivation, and self-efficacy) was found to be significantly associated with contextual factors, such as memory, metacognition, and determination strategies, and with the perception of Web-based collaboration in this study. It is postulated that students who are motivated to learn any given materials self-regulate their own learning and engage in cognitive processing, and in turn will achieve success in what they attempt. Such robust evidence has been found not only in traditional classroom-based studies (e.g., Pintrich & de Groot, 1990; Schmidt et al., 1996) but also in technology-based studies (e.g., Chang, 2005; Cho & Jonassen, 2009; Lee et al., 2005).

The social strategy was reported to be less frequently employed than other strategies, such as memory, and it was less likely to be associated with other factors in this study. These results parallel those of studies of L2 learning strategies (e.g., Bedell & Oxford, 1996). In contrast, such strategic processing (e.g., interacting with others and receiving support from peers and teachers) was found to play an essential role in learning an FL (e.g., Bown, 2009; Lai et al., 2012). There are several possible reasons for the differences between the findings of this study and those of previous studies. First, strategy use is influenced by multidimensional factors, such as culture, motivation/attitude, and willingness, as pointed out by Politzer and McGroarty (1985), who stated that Asian learners are unlikely to engage in socially strategic processing. Second, as their English teacher, one of the researchers in the present study found that most of the students were less willing to participate in collaborative work or group discussions in the classroom, and these informal classroom observations were consistent with the results obtained in the questionnaire survey.

The results of this study, which suggest positive attitudes toward learning in a Web-based collaborative environment, are in line with the findings of previous studies of the use of technology for learning beyond the classroom (e.g., Lai & Gu, 2011; Lai et al., 2012) and the associated enhancement of motivation and confidence (e.g., Kabilan et al., 2010). Such perceived usefulness and ease of use underscore the psychometric properties of goal-orientation, self-perceived abilities, and free agency of decision-making (Davis, 1989).

The evidence drawn from the current work suggests that it is essential for teachers to understand what types of vocabulary strategies their students tend to employ when learning new vocabulary words both inside and outside the classroom, and to further help students to regulate their own strategic processing by employing new strategies, such as social and metacognition strategies, which are effective but unfamiliar to them. Moreover, using Web-based tools—such as the one used in this study or others such as Facebook—is beneficial not only to support learning by students but also to increase their motivation and interests in the target language. However, several caveats still need to be addressed. First, this study examined the general aspects of SRL by the participants of this study rather than the details of how students regulate their learning in a given environment. Considerable efforts are still required to assess the development of self-regulated vocabulary strategy use in a longitudinal research paradigm. Second, how the participants performed in the learning tasks as part of a group collaboration might have been affected by their internal or emotional state; for example, participants who were shy or had a low proficiency level in English might not have felt comfortable providing answers to the given questions or commenting on the answers provided by others. Moreover, the classification of passive and active collaboration is limited to a single learning task that was chosen at random; thus, the classification is context-specific and subject to change according to the contents of the learning activities. In other words, the presence of less complex vocabulary in a given text might have encouraged more active collaboration, or vice versa. These aspects require further investigation. Finally, the difference in vocabulary improvements between those who tend to collaborate with each other and those who tend to work alone also needs to be explored further.
Acknowledgements

The authors would like to thank the Ministry of Science and Technology, Taiwan, R.O.C., for financially supporting this research under Grant Nos. NSC 101-2511-S-003-031-MY3, and MOST 103-2628-S-003-002-MY3 and MOST 103-2511-S-424-004-. They are also thankful that this research is partially supported by the “Aim for the Top University Project” and “Center of Learning Technology for Chinese” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C., the “International Research-Intensive Center of Excellence Program” of NTNU and the Ministry of Science and Technology, Taiwan, R.O.C. under Grant No. NSC 103-2911-I-003-301.

References


A Study of Learner-Oriented Negative Emotion Compensation in E-learning

Jiwei Qin1*, Qinghua Zheng2 and Haifei Li3

1Center of Network and Information Technology, Xinjiang University, Urumqi, China // 2Department of Computer Science and Technology, Xi’an Jiaotong University, Xi’an, China // 3Department of Computer Science, Union University, Jackson, TN, USA // jwqin.2008@gmail.com // qhzheng@mail.xjtu.edu.cn // hli@uu.edu

*Corresponding author

(Submitted May 14, 2013; Revised December 29, 2013; Accepted March 10, 2014)

ABSTRACT

E-learning provides an unprecedented flexibility and convenience for e-learners by breaking the limitations of space and time. However, the role of emotion is neglected in current e-learning systems. We focus strictly on negative emotions of e-learners, integrating emotion regulation theories with recommender technique, and present the study of learner-oriented negative emotion compensation in this paper. Inspiring from the existing emotion regulation model in the field of psychology, we design the architecture of learner-oriented negative emotion compensation in e-learning. Finding e-learner’s personalized emotion regulation strategies and methods from questionnaires, we propose an approach of e-learner’s negative emotion compensation based on recommender users and music. In practice, the usability of emotion compensation is verified by an e-learning platform, as a complementary demonstration, the learner satisfaction is done by a 90-e-learner survey in real e-learning. The results show that the proposed learner-oriented negative emotion compensation provides greater satisfaction for the e-learner, and is a feasible and effective method for e-learner to decrease negative emotions in e-learning.

Keywords

E-learning, Emotion compensation, Learner-centered design, Recommender system

Introduction

Emotion plays a significant role in the cognitive process of humankind. Neurobiology addresses the emotional basis of our full repertoire of cognitive options, including learning, attention, memory, and social functioning, all of which involve emotional process (Afzal & Robinson, 2010). Psychology also demonstrates that emotion is significantly related to students’ motivation, learning strategies, cognitive resources, self-regulation, and academic achievement (Pekrun et al., 2002). Empathy with the learner’s emotion would increase their motivation in learning (Pérez-Marin & Pascual-Nieto, 2012). All findings indicate that the formation of a complex cognitive system is usually accompanied by emotions.

Emotions of the user are usually classified into three categories: positive, neutral and negative. As a double-edged sword, emotions possess the dual use of either hindrance or help. Negative emotions turn an easy task into one difficult to perform, while positive emotions lead to an easier task performance. When negative emotions are too strong, performance is inhibited (John & Gross, 2004). In learning systems as well, optimal emotions can enhance enthusiasm for learning, but negative emotions suppress learners’ interest.

E-learning provides an unprecedented flexibility and convenience for learners by breaking the limitations of space-time. While most researchers are concerned about the learners’ cognitive process, they neglect the emotional role in current e-learning systems. The interactions of e-learners and teacher-learner rely on keyboard and mouse, both of which are “blind” (the visual function), “dumb” (the language capability), and “deaf” (the auditory function), without the least emotion. The e-learners can hardly feel emotional stimulation and in learning situations, we present the study that an learner survey in real e-learning process, so boredom arises and learning interest and learning efficiency are diminished. The inadequacy of affect is provoking great e-emotional interaction concern in recent years, and negative emotion compensation has emerged in the e-learning domain.

Specifically, focusing on negative emotions of e-learners in typical e-learning situations, we present the study that an effective approach is proposed to decrease e-learners’ negative emotions. The remainder of this paper is organized as follows. We introduce related work in Section 2. Next, after analyzing e-learners’ interactions in four modes of delivery, we propose the architecture of learner-oriented negative emotion compensation in e-learning, and describe
it in detail. Section 4 presents a prototype to check proposed approach’s feasibility and a 90-e-learner survey to check e-learner’s satisfaction on negative emotion compensation in e-learning. Finally, Section 5 discusses and concludes the paper.

Related work

Technology is used to implement different pedagogies, whether based on prevailing psychological conceptions, novel psychological explanations, or pedagogical justifications. They are tightly coupled and reciprocal (Salomon & Almog, 1998). In this section, we introduce the psychological concepts, including the definition and model of emotion regulation, and apply them for negative emotion compensation in e-learning.

Most research on emotion regulation has been done in the field of psychology, where psychologists have argued the definition of emotion regulation in much of literatures. Thompson (1994) states: “Emotion regulation consists of the extrinsic and intrinsic processes responsible for monitoring, evaluating, and modifying emotion reactions, especially their intensive and temporal features, to accomplish one’s goals.” He emphasizes that setting goals and taking the best opportunities are very important in emotion regulation. Gross (2001) describes: “Emotion regulation includes all of the conscious and non-conscious strategies we use to increase, maintain, or decrease one or more components of an emotional response.” Individuals may have different emotion regulation strategies and methods. Larsen (2000) proposes that in mood regulation the emphasis is more on mood per se, on altering the ongoing affective state without much reference to objective life events. Emotion regulation may be conceptualized as consisting of a series of distinct but interrelated control processes. The model is an application of a general cybernetic model involving feedback of negative emotion regulation.

The applications of emotions grew to cover a rage of techniques in learning. REP (Reengineering Educational Pedagogy based on conceptualizing the impact of emotions upon learning) is the MIT research project that aims to establish an executable learning partner which can track the learners’ emotion state in learning process (Kort et al., 2001). The REP model depends on Plutchik & Ekman's basic expression to establish an emotion model for learners in the learning process. Based on the REP model, the integration of characteristics and goals determines the learners’ emotion state (Shen et al., 2009). In computer adaptive tests, adaptive feedback of emotions is characterized by negative emotions to describe emotional behaviors (Economides, 2006). A negative anxiety-frustration loop dominates learner’s emotions, making it necessary to break the negative loop by means of psychological theories and technology (Juutinen & Saariluoma, 2012). Research results help us to know the fundamental characteristics of learners’ emotion regulation, and inspire us to design an appropriate architecture of negative emotion compensation for e-learners.

The architecture of negative emotion compensation

Breaking down the e-learners’ behavior into the four modes of delivery (i.e., real-time classroom, group discussion, courseware on demand and questioning-answering), and inspiring from the existing emotion regulation model (Larsen, 2000; Gross, 2001) in psychology domain, we define the architecture of learner-oriented negative emotion compensation in e-learning. The architecture consists of three modules: emotion recognition, personalized emotion regulation and negative emotion compensation, and is shown in Figure 1. First, we recognize e-learner’s current emotion state. If an e-learner has some optimal emotions (or neutral calmness), she go on with her study in e-learning. Otherwise, she is redirected to the personalized emotion regulation module. Then the compensation list (music and user) is presented to the e-learner based on her historical compensation cases until the e-learner’s negative emotion is alleviated to pursue her study.

The detailed analysis of the learner-oriented negative emotion compensation is demonstrated in following section. In general, there are four modes of delivery in e-learning: real-time classroom, group discussion, courseware on demand and questioning-answering. The real-time classroom resembles the real physical classroom in that it can elicit learners’ emotions. Group discussion presents the usual learner-to-learner interaction pattern similar to a chat room, where users’ communication mainly depends on interactive text. And almost all of the information exchange between e-teachers and e-learners are lost in courseware on demand. Finally, Q&A, similar to the BBS, also relies on interactive text, either online or off-line. The interaction patterns of these can be classified into synchronous.
interaction and asynchronous interaction. The former includes real-time classroom, group discussion and online Q&A, and the latter contains courseware on demand and off-line Q&A.

![Figure 1. Learner-oriented architecture of emotion compensation](image)

The emotion recognition is regarded as the prerequisite of the negative emotion compensation. E-learners’ emotions are recognized by multimodal communication, including facial expression, speech, body language, behaviors and interactive text. The recognition from facial expression (Gnjatovic & Rosner, 2010), speech (Pantic & Rothkrantz, 2000) and body language (Schindler et al., 2008) requires learner to be equipped with some detection devices, such as wearable sensors and high resolution cameras. However, in real scenarios, e-teacher often face with thousands of online e-learners, it is almost impossible for each e-learner to be equipped with expensive video and audio analysis devices with the limitations of costs and network bandwidth, especially for the schools or families in developing countries, such as China (Tian et al., 2009). So, we emphasize the emotion recognition from interactive text and behaviors from the e-learners’ logs in asynchronous interaction in this study.

**Current emotion recognition**

*Interactive text*

Interactive text is the most popular communication means in modes of delivery, such as group discussion, courseware on demand and questioning-answering. Because the interactive text is composed of some sentences, the emotion recognition from interaction text depends on the analysis of sentences’ sentiment. Between the syntactical rules and affective lexicon are the fundamental factors in the analysis of sentences’ sentiment at interactive text. Supposed the sentence is a compound sentence, conjunction should be concerned. If it is successive (e.g., and, or, and so etc.), the resulting vector is the maximum intensity in vectors of both clauses; if it is adversative (e.g., but,), the resulting vector is following after the connectors clause (Neviarouskaya et al., 2007). Or else, this sentence is simple sentence. Following, the word device is used to selected emotion words from a simple sentence (or a clause). For each emotion word, its affective feathers are described by a vector \( e = (\lambda_1, \ldots, \lambda_m) \), \( m \) is the category of basic emotions. Variable \( \lambda \) is the set of intensities corresponding with basic emotions. If \( \lambda \) represents the intensity of positive basic emotion, \( \lambda_i \) is from 0 to 1, ibid, if \( \lambda_i \) represents the intensity of negative basic emotion, \( \lambda_i \) is from -1 to 0. The sentiment of sentence is expressed by equation (1), here \( m \) refers to the number of emotion words, according to different modifier (e.g., very, extremely and most etc) and prefixes (e.g. un and non), \( \sigma \) is the weighted value.
\[ E = \sum_{j=1}^{m} a_j e_j \]  

When \( E \) is equal to zero denotes that the emotional value of sentence is zero, the learner’s current emotion is calmness; When \( E \) is greater than zero means that the emotional value of sentence is positive number, the learner’s current emotion is the positive; On the contrary \( E \) is negative number, the e-learner’s current mood is negative. Of all these three emotional states of e-learner, negative emotions should be regulated in the study.

**Behavior in courseware on demand**

The e-learner’s behaviors in courseware on demand come from the log of Internet Education School of Xi’an Jiaotong University (http://www.dlc.xjtu.edu.cn/) from May 2010 to September 2012. The number of CoD (Course on Demand) in a day is illuminated in Figure 2. It can be seen that the e-learners’ demand for CoD varies greatly. The playing time of network educational resource files is about 49 minutes, supposing that the unit time interval is three minutes, the possible distribution of the cumulative probability of stop-on-demand in a file of CoD is described by Figure 3, the transverse coordinates indicate the length of playing time, and longitudinal coordinates show the cumulative probability of stop-on-demand. Thus from the above in-depth analysis, we arrive at the following views: when e-learners demand the courseware, they wait too long on courseware due to the limitations of service and network. As a result, e-learners lose patience and stop on demand; when e-learners roughly understand the content of the courseware, after watching for a few minutes, the content of the courseware causes e-learners to lose interest in learning; for popular courseware, e-learners may have a herd mentality leading to more e-learner requests on-demand. When e-learners start a video which is not the desired courseware, they stop playing.

![Figure 2. The number of CoD in a day](image)

![Figure 3. The cumulative probability of stop-on-demand](image)
Personalized regulative strategy of emotion

Emotion regulative strategy not only provides powerful theoretic instruction for emotion compensation, but also acts as a bridge between recognition and compensation. It is a key element of the learner-oriented architecture of emotion compensation. Because individuals have difference in personality traits, they may use different emotion regulative strategies in e-learning. Integrated emotion regulative strategies with the personality of e-learners, the corresponding emotion regulative methods of e-learners are determined.

Questionnaire is often used to construct and verify theories in the domain of pedagogy and psychology (Gross, 2001). We intend to use three questionnaires (the Big Five Questionnaire, ERQ and NMR-S) to acquire e-learners attributes including personality, emotion regulation strategy and emotion regulative method and construct the matching rule base for emotion compensation.

In the present study, the Chinese version of questionnaires (ERQ and the Big Five) are developed with a back-translation procedure and used to measure emotion recognition strategy and personality of participants. One bilingual Chinese-English person translated the English version of the ERQ and the Big Five into Chinese. Discrepancies emerging from this back-translation were discussed and adjustments to the Chinese translation of the ERQ and the Big Five were made.

The Big Five depends on a rapid detection version by Shafer (1999), which is a 30-item short form and consist of five dimensions, neuroticism (N), extraversion (E), open (O), Agreeableness (A) and conscientiousness (C), each dimension contains six items which are a pair of bipolar adjectives. And ERQ is a 10-item questionnaire, which is designed to assess individual differences in the habitual use of two emotion regulation strategies: cognitive reappraisal and expressive suppression. Emotion regulation strategies (reappraisal and suppression) questionnaire which is comprised of 10 items, six items of them are used to appraise the cognitive reappraisal, the others are used to appraise the suppression expression. The items are rated on 7-point scale ranging from 1 to 7 scales, with 1 being strongly disagree and 7 being strong agreed. The data of e-learners’ personality and emotion regulative strategy are recorded and used in negative emotion compensation module.

A 20-item questionnaire named the Negative Mood Regulative Scale which is operated under the assumption that users encounter the predicament of emotion in the e-learning process. Every given item indicates a concrete action, for example: “You would like to chat with friends?” Respondents select a value from 1 to 5 where a high value represents strong agreement, and a low value represents strong disagreement.

In order to obtain reliable data of samples, we select participants that have an experience in e-learning. A total of 60 participants are recruited through the BBS of Xian Jiao tong University in Chinese. They haven’t a history of psychopathology and neurological impairments via an open-ended format in their self-report on a demographic questionnaire. To ensure the quality of data, we give participants brief oral instructions including the importance of real data and the objective of investigation. Participants can cope with the study via paper-based questionnaires and e-questionnaires in Internet. The results show that 97% of respondents believe doing something is helpful to relieve negative emotional tension, about 95% of respondents approve of enjoying their favorite activities (e.g., music, movie) and 90% prefer chatting with friends to share their emotional predicament.

From above results, learner would like to listen to music in negative emotional state. Music loads emotional content to reduce the listener’s current emotion (Kaminskas & Ricci, 2012). Listeners can perceive and produce an emotion when music reaches their ears (Juslin & Sloboda, 2001). Listening music as a popular way can evoke powerful emotions and compensate negative emotion of e-learners.

Negative emotion compensation

Among the four modes of delivery, face-to-face interacting with teacher is done in the real-time classroom; e-teacher and peer assist e-learner to relieve negative emotion in group discussion and online Q&A; and expert’s guiding paired with music help e-learners to cope with negative emotion in CoD and off-line Q&A. As is mentioned above, we present an algorithm of negative emotion compensation to automatically recommend music and users (expert, e-teacher and peer) to alleviate e-learners’ negative emotion.
The process of negative emotion compensation

The result has proved that personality traits relate to emotion regulating (Qin et al., 2011). Extroverted personalities prefer to select cognitive reappraisal to regulate negative emotion, whereas neurotic types are always apt to select suppression of expression (Dennis, 2007). We can anticipate e-learners’ emotion reaction based on personality (Hoerger & Quirk, 2010). Moreover, several studies touching upon trust and personality traits show that people with similar personalities are more trustful (Golbeck, 2009) of one another. We associate emotion with trust and personality traits in negative emotion compensation. The process of negative emotion compensation is divided into three steps, which includes computing similarity score, predicting ratings and producing compensation list. It is illustrated in Figure 4.

**Figure 4.** The process of negative emotion compensation

The similarity score is used to measure the preference relation among e-learners. The larger score of the similarity means the more similar between e-learners. Several ways have been used to measure the user similarity score, such as Pearson correlation, Spearman correlation and Cosine correlation. Comparing the existing experimental result, the Pearson correlation is better than the Cosine correlation on the performance of calculating similarity score (Breese et al., 1998) by ratings. However, The Spearman correlation computes a measure of correlation between ranks instead of ratings (Herlocker et al., 2002). So the Pearson correlation is used to measure e-learner similarity on ratings. The evidence (Golbeck, 2009) shows that the higher the trust value between two users, the larger the similarity. So the rating value and the trust value jointly decide the similarity score in emotion compensation. Meanwhile, when insufficient or even no items are co-rated by e-learners, the trust value is used as the similarity score. The trust value of fresh e-learner can be evaluated by personality from the personalized emotion regulation module. The similarity score is expressed in equation (2).

\[
\omega_{u,v} = \kappa p_{u,v} + (1-\kappa)t_{u,v}
\]  

(2)

\[
p_{u,v} = \frac{\sum_{i \in I_{u,v}} (R_{u,i} - \overline{R_u}) \cdot (R_{v,i} - \overline{R_v})}{\sqrt{\sum_{i \in I_{u,v}} (R_{u,i} - \overline{R_u})^2} \cdot \sqrt{\sum_{i \in I_{u,v}} (R_{v,i} - \overline{R_v})^2}}
\]  

(3)

Where \(\omega_{u,v}\) is the similarity score between learner \(u\) and learner \(v\), \(t_{u,v}\) is the trust value between learner \(u\) and learner \(v\), \(p_{u,v}\) is the similarity score on ratings given by learner \(u\) and learner \(v\), and is calculated by expression (3). Here \(R_{u,i}\) is rating that learner \(u\) rated item \(i\) (i.e. music), \(\overline{R_u}\) is the average rating of learner \(u\), \(I_{u,v}\) are the co-rated items by learner \(u\) and learner \(v\). The rating value of unknown item is calculated by equation (4).

\[
R_{u,i} = \overline{R_u} + \frac{\sum_{v \in I_{u,v}} \omega_{u,v} \cdot (R_{v,i} - \overline{R_v})}{\sum_{v \in I_{u,v}} \omega_{u,v}}
\]  

(4)
Where $U$ is the set of learners and $j$ is an unknown item of learner $u$. The compensation list of top $n$ musical tracks and $m$ users are recommended to compensate negative emotion for the e-learner.

Besides, when a fresh e-learner enters emotion compensation system, she has no historical ratings and trust relations with others. To solve the fresh e-learner’s problem, we can find similar e-learners based on her personality traits and emotion regulation strategy.

**The algorithm of negative emotion compensation**

The algorithm of negative emotion compensation is presented in table 1.

<table>
<thead>
<tr>
<th>Table 1. The algorithm of negative emotion compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong></td>
</tr>
<tr>
<td>$T$, e-learner’s trust network;</td>
</tr>
<tr>
<td>$U$, the set of users;</td>
</tr>
<tr>
<td>$A$, the ratings of e-learners;</td>
</tr>
<tr>
<td>$M$, the set of music;</td>
</tr>
<tr>
<td>$u$, an active e-learner;</td>
</tr>
<tr>
<td>$e$, the emotion state of $u$.</td>
</tr>
<tr>
<td><strong>Output:</strong></td>
</tr>
<tr>
<td>$L$, a list of music and users.</td>
</tr>
<tr>
<td><strong>Steps:</strong></td>
</tr>
<tr>
<td>1. Set the $u$ as the root of $T$ and $u \in U$;</td>
</tr>
<tr>
<td>2. Select ratings which are rated together by $u$ and others under $e$;</td>
</tr>
<tr>
<td>3. Compute the similarity score between $u$ and $v$ on the ratings;</td>
</tr>
<tr>
<td>4. Compute the similarity score between $u$ and $v$ based on $\alpha_{u,v} = \kappa \rho_{u,v} + (1-\kappa)t_{u,v}$;</td>
</tr>
<tr>
<td>5. Select top $m$ neighbors of $u$ according to $t_{u,v}$;</td>
</tr>
<tr>
<td>6. Run $R_{u,j} = R_u + \frac{\sum_{v \in U} \alpha_{u,v} \left( (R_{u,j} - R_v) \right) \alpha_{u,v}}{\sum_{v \in U} \alpha_{u,v}}$;</td>
</tr>
<tr>
<td>7. Find top $n$ pieces of music;</td>
</tr>
<tr>
<td>8. Return $L$.</td>
</tr>
</tbody>
</table>

**Experiments**

In order to demonstrate the performance of the algorithm of negative emotion compensation, we compared the other two learner’s preference measures respectively using the correlation on the co-ratings and the trust value on the different sparsity of the rating matrices of datasets, Epinions and EC. The correlation on the co-ratings is used to weight user’s similarity and judge the user’s preference in the collaborative filtering (CF for short), while the trust value is regard as user’s similarity in the trust-based recommendation (TR for short). The learner’s preference is judged by the learner’s similarity, which is calculated selectively using the correlation on the co-ratings and the trust value in negative emotion compensation (NEC for short). The experimental results show that NEC outperforms CF and TR, which provides a basis for implementation of emotion compensation, and learner satisfaction increases in the learner-oriented negative emotion compensation system.

**Dataset**

There are several datasets which are usually used to do experiments on a trust-based recommendation, in particular one named Epinions is extracted from the trust network datasets released at trustlet.org (Massa & Avesani, 2006), which was derived from Epinions.com by Paolo Massa and Paolo Avesani. In the dataset, the ratings follow the 1, 2, 3, 4, and 5 numerical scales, the trust value is either the value of 0 or the value of 1; obviously, the value of 0
indicates the trust relationship between user A and user B does not exist, while the value of 1 represents that user A trusts user B. There are 49,290 users and 139,738 items in the dataset from Epinions.com, where each user rated at least one item in the past. The sparsity of the rating matrices of Epinions is 99.99135%.

Because of the lack of reference data and because of the limitation of unified categories of emotions (Kaminskas & Ricci, 2012), there is no popular dataset for NEC. Following the principle that the categories of emotions are determined by the task and domain of research, we consider a scenario where we use music to regulate a learner’s negative emotions when they encounter the predicament of emotion in the e-learning process. These negative emotions figured frequently in e-learning: anxiety, anger, disgust, sadness, shame, and hopelessness (Tian et al., 2009). To illustrate and validate our work, we endeavored to set up a website to acquire data, named “EC” in our experiment. Volunteers were recruited from Xi’an Jiaotong University in China, and 102 were selected to rate music at a resource-making website. Participants did not have a history of psychopathology or neurological impairments, as self-reported via an open-ended demographic questionnaire. In this experiment, we collect 1548 pieces of music taking into account the multidimensional features of music as well as cross-cultural differences. Each piece of music is rated in different negative emotional contexts. All ratings follow the 1-extremely bad, 2-bad, 3-average, 4-good, and 5-perfect numerical scale. Participants are expected to annotate randomly selected music for eight weeks. While listening, they can click on the left rectangle to pause/play the song; drag-and-drop the middle rectangle to listen to the song again; drag-and-drop the right rectangle to adjust the volume; and click on the button to modify the annotation. Further, titles are shown on the webpage. In total, 102 participants gave 21,738 ratings. The sparsity of the rating matrices of EC is 96.663%.

Evaluation

In general, both accuracy and coverage are usually used to evaluate the performance of recommendation algorithm (Herlocker et al., 2004; Ruffo & Schifanella, 2009; Herlocker et al., 2002). The accuracy is measured according to Mean Absolute Error (MAE), Mean Absolute User Error (MAUE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). If \( \{p_1, p_2, \ldots, p_n\} \) refers to the predicted ratings, and \( \{r_1, r_2, \ldots, r_n\} \) is the actual rating, then the MAE, MSE and RMSE are expressed by equation (5), (6) and (7), respectively.

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |p_i - r_i|
\]

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (p_i - r_i)^2
\]

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - r_i)^2}
\]

As mentioned above, suppose \( U \) is the set of all users and \( m \) refers to the number of users, \( h \) is the number of ratings given by \( U \), then the MAUE and MAUE_0 are expressed by equation (8) and (9).

\[
\text{MAUE} = \frac{1}{m} \sum_{i \in U} \text{MAUE}_i
\]

\[
\text{MAUE}_i = \frac{1}{h} \sum_{i \in U} |p_i - r_i|
\]

The coverage is a measure of the percentage of predictions given by the recommendation algorithm, and includes rating coverage (RC) and user coverage (UC), which are respectively denoted as equation (10) and (11).

\[
\text{RatingCoverage} = \frac{p}{s}
\]

\[
\text{UserCoverage} = \frac{k}{m}
\]

where \( s \) is total number of all ratings, \( p \) is the quantity of ratings predicted, \( m \) is the total number of all learners, and \( k \) is the number of learners predicted.

Because emotion is a subjective psychological experience, self-reporting is an important evaluation tool originating from feedback in the psychological domain. We measure whether the learners consider the music recommended by NEC is best satisfying than those by CF, TR. In order to perform this evaluation of NEC, the learners’ satisfaction degree (SD for short) is proposed and expressed by equation (12).
\[ SD = \sum_{i=1}^{5} r_i \times \frac{1}{i \times m} \]  

(12)

where \( r_i \) represents ratings on a scale from 1 to 5, with the higher value representing more satisfaction, \( r_i \) is the quantity of \( r_i \), \( i \) is the quantity of recommended music for one learner.

**Experimental results**

We run NEC, CF and TR on Epinions, the accuracy and coverage are shown in Figure 5 and Figure 6, respectively. It is seen that the MAE, MAUE, MSE, RMSE of NEC is lower than that of CF and higher than that of TR; the rating coverage of NEC is much higher than that of CF and TR, respectively; and the user coverage of NEC is much higher than CF and TR.

**Figure 5.** The accuracy of NEC, TR, CF on Epinions  
**Figure 6.** The coverage of NEC, TR, CF on Epinions

Next, Figure 7 indicates that NEC and CF hardly differ in the MAE, MAUE, MSE, RMSE, which both offer better accuracy than TR; and the rating coverage of NEC and TR is much higher than that of CF, even attaining 100% on user coverage in Figure 8. From the above experimental results, the NEC has better performance on datasets with different sparsity of the ratings matrix.

**Figure 7.** The accuracy of NEC, TR, CF on Epinions  
**Figure 8.** The coverage of NEC, TR, CF on Epinions

Furthermore, Table 2 illustrates the result of learners’ subjective evaluation under an emotional context of anxiety: NEC gains the better satisfaction of recommendation. In a word, NEC shows better performance on learner’s satisfaction.

**Table 2.** Learner’s satisfaction degree of CF, TR and NEC in anxiety

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>The number of ratings</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>3.656</td>
</tr>
<tr>
<td>TR</td>
<td>1</td>
<td>9</td>
<td>27</td>
<td>46</td>
<td>17</td>
<td>17</td>
<td>3.644</td>
</tr>
<tr>
<td>NEC</td>
<td>1</td>
<td>8</td>
<td>29</td>
<td>35</td>
<td>17</td>
<td>17</td>
<td>3.656</td>
</tr>
</tbody>
</table>
Visualization of negative emotion compensation

The emotion compensation system as a prototype has been designed to verify the availability of the architecture of negative emotion compensation. And the mirror of learner's satisfaction of negative emotion compensation is reported through the system.

When a fresh e-learner logins emotion compensation system, she would complete three questionnaires (Big Five, ERQ and NMR-S). Following, if the e-learner encounters difficulties in e-learning process, she may raise negative emotion that is recognized immediately. Then the system auto-recommends appropriate music and users (peer, teacher and expert) to the e-learner based on trust and historical compensation data. For example, suppose an e-learner is talking about the topic of final examination of College English, when “I’m very anxious” appeared in interactive text, the sentence is analyzed. Because the phrase “very anxious” is the dominant emotion of the sentence, and modifier “very” increases the intensity “anxious.” The output result displays that this sentence expresses negative sentiment, suggesting “my” emotions should be compensated. Five pieces of music and users are present to e-learner. It is shown in Figure 9.

Figure 9. A prototype: Negative emotion compensation system

In this paper, a 90-e-learner survey is used to check e-learner’s satisfaction of emotion compensation with a list of music and users under negative emotional situation. The evaluation results of e-learners are presented in Figure 10. The vast majority of e-learners are satisfied with the compensation list of music and users. We verify that the learner-oriented emotion compensation works well enough to alleviate e-learner’s negative emotions in e-learning.

Figure 10. E-learner’s evaluation
Conclusions

This paper shows a study of learner-oriented negative emotion compensation in e-learning. Inspiring from existing emotion regulation model in psycho, we design the architecture of learner-oriented negative emotion compensation. Further analysis of the e-learners’ behaviors in modes of delivery, we emphasize the emotion recognition of interactive text and the behaviors in CoD from the log of the Internet Education School of Xi’an Jiaotong University. In order to find e-learner’s personalized emotion regulative methods, three questionnaires (the Big Five, ERQ and NMR-S) are adopted. Integrating the results of surveys and existing technique, we propose the approach of e-learners’ negative emotion compensation based on recommender music and users. Furthermore, emotion as a subjective experience, a comprehensive measure is presented to evaluate the proposed approach. Both MAE and coverage are used to evaluate the performance of algorithm. On the other hand, a prototype system is created as a negative emotion compensation application to verify the usability of architecture of negative emotion compensation. A promising result is shown by a 90-e-learner survey, the proposed approach is practical and effective for alleviating e-learners’ negative emotion in e-learning.

From the analysis of e-learners’ behaviors in CoD, we found that both the content of the courseware and the limitations of service and network can cause the e-learner’s emotion change. It is suggestion that the construction and management of learning resource is as important as emotion interaction in e-learning.

Moreover, because emotion is complex, there is not universal taxonomy of emotion. Focusing on the goal and domain of research, researchers are to choose different lists of emotion (Kaminskas & Ricci, 2012). The study refers only to the negative emotions of anxiety, sadness, hopelessness, anger, shame and disgust in the learning domain. This is only because above six negative emotions take place most often in e-learning. However, negative emotions are not limited to these in learning and other domains. Doubtless, the proposed architecture and algorithm can be implemented to regulate e-learners’ negative emotions.

It is worthwhile to mention that the recommendation is not the only technique for negative emotion compensation. With the development of technology, we will find the suitable technique to absorb in negative emotion compensation. It is a great challenge to simulate realism of negative emotion compensation in e-learning.

Acknowledgements

The research was supported in part by the National Science Foundation of China under Grant Nos. 91118005, 91218301, 41261087, 61308120, National High Technology Research and Development Program 863 of China under Grant No. 2012AA011003, Key Projects in the National Science & Technology Pillar Program under Grant No. 2013BAK09B01, Cheung Kong Scholar’s Program. We would like to thank the anonymous reviewer for insightful comments.

References


431
Recognition-Based Physical Response to Facilitate EFL Learning

Wu-Yuin Hwang¹, Timothy K. Shih², Shih-Ching Yeh²*, Ke-Chien Chou¹, Zhao-Heng Ma²
and Worapot Sommool³

¹Graduate Institute of Network Learning Technology, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan // ²Computer Science and Information Engineering, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan // ³Department of Computer Engineering, Kasetsart Rajabhat University, Thailand // wyhwang@cc.nchu.edu.tw // timothykshih@gmail.com // shihchihy@csie.nchu.edu.tw // maplewind127@gmail.com // maxma1021@gmail.com // worapot@fci.ksu.ac.th

*Corresponding author

(Submitted October 07, 2013; Revised December 13, 2013; Accepted January 22, 2014)

ABSTRACT

This study, based on total physical response and cognitive psychology, proposed a Kinesthetic English Learning System (KELS), which utilized Microsoft’s Kinect technology to build kinesthetic interaction with life-related contexts in English. A subject test with 39 tenth-grade students was conducted following empirical research method in order to compare the effectiveness of KELS on both learning and motivation. In addition, we developed one questionnaire to study the perception learning style of the students involved in the experiment in order to determine whether students with different perception styles experienced any difference in learning while using the proposed KELS. Finally, we investigated students’ acceptance of and motivation toward using KELS. The result showed that there was a significant difference between the control group and the experimental group regarding long-term retention. Moreover, the experimental group’s learning was significantly affected by using kinesthetic interaction between peers to facilitate learning to speak and listen in English. The study also found that KELS was effective across perceptual learning styles. Finally, most participants in the experimental group agreed that KELS would effectively increase their motivation to learn English and expressed a strong intention to continue using the system.

Keyword
Learning, Kinesthetic, Language, Interaction, Perception

Introduction

The importance of English proficiency grows in the trend of globalization. Learning EFL (English as a foreign language) in Asia Pacific, however, particularly in Taiwan, is still mostly for the purposes of enrolling in higher education institutions; thus, the examinations pay the most attention to reading, spelling, grammar and writing skills while usually neglecting daily life communication like speaking and listening. The development of long-term retention is necessary in order for EFL learners to be able to retrieve their knowledge of the English language to communicate with others or apply to daily life, although the ability to retain knowledge decreases over time (Ebbinghaus, 1885). EFL learners in Taiwan learn English beginning in elementary school and continuing to college without the benefit of having a supportive context for the language. Because the schools use traditional teaching methods, EFL learners still view English as necessary only for the purposes of examination rather than as an important communication skill or ability that relates to their daily lives or careers. This phenomenon violates the theory of situated learning, proposed by Suchman (1987), which emphasizes that learning occurs in a cultural, practical and meaningful context; knowledge cannot be separated from its context. For this reason, it is important to design and provide English activities for EFL learners to use that involve life-related contexts (e.g., conversation, presentation, or question-and-answer).

In recent years, a variety of teaching methods and technology to promote EFL learning has been widely discussed. For example, research supports the theory that physical movement can enhance the process of learning because involving learners’ interaction by gesture has a positive effect on increasing learners’ attention. Bruner (1996) proposed the theory of systems of representation, which states that the first stage of the learning cognition process is to enact learners to do what they learn, that is, to involve an active representation. Helping learners make physical motions in order to understand what they learned and interact with their surrounding environment improves learning by applying knowledge to related situations. Similarly, Gardner (1989)'s theory of multiple intelligences suggests a connection between learning activities and kinesthetic intelligence, that is, students learn and solve problems using
physical motion. One of the most well-known teaching methods connected with interaction between limbs is total physical response (TPR), which is useful for language learning as proposed by James Asher (1966). Using body motion or behavior to illustrate listening and understanding is one of the key concepts in TPR. Asher proposed that learners respond to auditory stimulus with body motion (e.g., nodding, shaking hands, and waving) not only to demonstrate their ability to listen well but also to help internalize what they learned deeply in order to improve and sustain the effect of their learning. Moreover, experts of brain science have great esteem for TPR. Body motion as a medium for learning can help to create a strong association between body motion and language, which improves their auditory learning skills.

Psychological factors play an important role in language learning as well. Krashen (1981) proposed the affective filter hypothesis, which posits that a good language learning environment must allow learners to be confident and relaxed and to try and fail without pressure and anxiety. Pressure and fear of failure seriously impede language learning. So TPR allows language learners at the beginning of their study to present the meaning of what they hear by using their bodies rather than requiring them to speak. In recent years, the application of the cognition learning style to language learning has gradually been noticed. Dunn (1983) classified four cognition learning styles: visual, auditory, kinesthetic, and tactile (VAKT).

Literature review

Multiple representations and interaction

Bruner’s (1996) system of representation theory proposes that the cognitive process of learning converted the exterior environment to inner development. The theory contains three periods of learning: enactive representation, iconic representation, and symbolic representation. The task in the first stage is to understand the exterior environment and events by enacting their motion. In the next stage, the learner converts perceptions received from the previous stage into iconic memory to explain the exterior events in order to obtain knowledge. In the last stage, learners convert symbols into meaningful and thoughtful symbol memory to explain what was perceived through the exterior environment and events.

Gardner (1993) proposed the theory that multiple intelligences are tools that people utilize in order to learn and solve problems. He cited eight types of intelligence: linguistic, logical, spatial, kinesthetic, musical, inter-personal, intra-personal, and naturalist. Since traditional teaching methods put emphasis on logical and linguistic types of intelligence, the possible benefits of adding methods that emphasize kinesthetic intelligence are lacking.

Combining representation theory with an emphasis on enactive representation and kinesthetic intelligence, this study designed four kinds of body-interactive mechanisms to facilitate learning EFL.

Information processing and learning gain

The learning retention effect is the remaining memory volume of learning content as measured a while after the learning process has ended without providing any further learning activity. Hermann Ebbinghaus (1885) found that the loss of memory coincided with the passage of time. Ebbinghaus’ forgetting curve showed that the volume of learning content decreased regularly, and the decreasing rate was highest at the beginning stage while memory was being generated. Over time, the decreasing rate lowered. Further, while learners performed a better comprehension with knowledge, the speed of forgetting went slower. If learners were forced to recite knowledge in a short time period without much comprehension, then the learning effect was negative, and the rate of loss of memory was high. Aside from the testing conducted after the experiment ended, this study performed a delay test on all participants 21 days following the conclusion of the experimental teaching in order to determine any differences between the control group and the experimental group regarding learning retention effect.
Total physical response with context support for language learning

Total physical response, proposed by James Asher (1966), is a language teaching method. Asher believed that the process of learning a foreign language is similar to that of learning a native language. To learn speaking naturally, children must experience a silent period and is considered as the readiness for production. At this stage, while they are not able to speak naturally, they repeatedly receive auditory stimulation from their family members or other people in their environment through bodily motions such as nodding, shaking the head, or waving. The interaction between listening and body motion can enhance the comprehension and internalization of language input and facilitate long-term retention.

The most important concept of the TPR teaching method is to have learners perform motions with their bodies to express their understanding of what they heard. The teaching process contains a series of motion commands; after listening to and understanding the meanings of words, learners make the corresponding motion to further perceive the meanings of the words through bodily sensation. The interaction between visual, auditory and motor perceptions brings about a better understanding of language. Once learners are able to handle voice factors well, speaking skills should be developed naturally (Richard & Rodger, 1996). Asher also suggested that not asking learners to develop speaking skills at the beginning stage of language learning can effectively reduce anxiety and improve learning.

Suchman (1987) found that learning is context-related and thus cannot be separated from the learners’ context. The acquiring of knowledge is more effective if learning activities are correlated with life tasks and have some real “meaning” to the lives of the learners. In view of these facts, this study proposed to utilize TPR and kinesthetic intelligence to bring together learning activities and meaningful life situations or contexts in order to improve the effectiveness of learning EFL.

Along with the creation of gesture-based interfaces, such as Microsoft Kinect or Nintendo Wii, educators have more opportunities to provide students more convenient way interacting with multimedia learning environments (Johnson, Smith, Willis, Levine & Haywood, 2011). Chang, Chien, Chiang, Lin and Lai (2013) suggested the embodied approach using Kinect was very promising to facilitate students’ cognitive learning outcomes in multimedia learning environments. Chao, Huang, Fang and Chen (2013) utilized Kinect as body-motion interfaces to develop embodied tasks in learning process. The results indicated that embodied tasks did help learners better memorize or recall learning contents in learning process.

Even thought TPR was proposed for language learning in some of the related work mentioned above, there is no interactive digital-learning system presented in these researches. Also, while Kinect was applied in some other related work, they were not specifically for language learning. As a result, this study was the pioneer to develop a Kinect-based interactive system specifically for EFL with which TPR was applied. Furthermore, this study also tried to investigate the behaviors of using our proposed Kinect-based interactive system deeply and their effects on EFL learning.

Learning motivation and perception style

The affective filter theory suggests that the attitude of learners affects the learning of a foreign language. For example, EFL learners acquired knowledge more effectively when depression and anxiety were reduced and when they were more positively motivated to learn. Hence, motivating the learner is a very important element of any teaching strategy (Keller, 1987). Keller argued that the traditional teaching design pays less attention to the motivation of learners and thus reduces the effectiveness of learning. To increase motivation, he integrated various theories of learning psychology, including attribution, achievement motivation, and expected value, and proposed the ARCS (Attention, Relevance, Confidence, Satisfaction) model of motivation to address relationships between learning motivation, teaching material design, and learning effect. He also reinforced the systematic teaching material design in order to facilitate learners to participate and interact with learning activities. In this study, we adopted Keller’s Learning Motivation ARCS Questionnaire to test whether the KELS (Kinesthetic English Learning System) can increase learners’ motivation to learn EFL.
A perception style is defined by the way people prefer to process information and by the information type people process skillfully. In the past, researchers paid more attention to the interaction between cognition style, emotional factors and situation demands while studying EFL (Brown, 1974; Ely, 1986; Hatch, 1974; Heyde, 1977; Naiman, Frohlich, & Todesco, 1975; Tarone, Swain, & Fathman, 1976; Tucker, Hamayan, & Genesee, 1976). In his research with American children, Dunn (1983) proposed the VAKT perception style classification: visual, auditory, kinesthetic, and tactile. Visual learners prefer reading text and viewing charts while auditory learners prefer listening to speech or audiotapes. Kinesthetic learners, on the other hand, prefer using body-interactive learning methods to experience knowledge, and tactile learners prefer the learning method that allows them to interact with the learning material through activities such as building models or performing laboratory experiments. Reid (1984) further applied the VAKT perception styles to the research on EFL learners and discovered that most students can correctly choose the style in which they prefer to learn. He also recognized that more EFL learners prefer kinesthetic and tactile learning methods as well as the peer learning model. Also, Lien (2012) proposed an interactive English learning system with Microsoft Kinect and examined what kind of learning style of students could benefit from it. Results showed that the motion-based system could significantly enhance student’s metacognition on reflective, sensing and sequential styles.

System design

Overview

There are two main modules in the proposed KELS: the activity module and the interactive mechanism. The activity module is composed of four learning activities: vocabulary learning, a listening test, a paired interactive speaking activity, and motion-based sentence-making. The interactive mechanism includes four types of interactions: mimicking motions, object recognition, gesture-based relative positioning, and kinesthetic clicking. The system design is shown in Fig. 1.

![Figure 1. System functional architecture](image)
Body interactive mechanism

Following the four types of interaction mechanisms, learners were able to interact with the learning system and receive information about the vocabulary (e.g., spelling, phonogram, sample sentence or pronunciation) by displaying it on a screen, as shown in Fig. 3.

Motion mimic

This mechanism allowed learners to acquire motion- and pose-related vocabulary. Learners were instructed to make the corresponding pose to demonstrate the meaning of the vocabulary word in order to trigger the display of information about the word. We expected that learners would better understand and internalize the vocabulary by interpreting the meaning of the vocabulary through the performing of motions and poses.

Object recognition

When learners intended to learn the vocabulary of a specific object in their environment, they only needed to pick up the object in order for the KELS to recognize it and provide information about the corresponding vocabulary. The act of picking up an object was expected to enhance the learners’ impression and perception of the vocabulary.

Gesture-based relative positioning

This mechanism was for learners to understand a series of prepositions indicating relative position between objects. When learners put their arms out straight in order to point to a specific orientation, the KELS automatically detected the preposition suitably expressing the relative position between the learners’ location and the orientation the learners were pointing to and displayed information about the preposition. Learners were expected to regard this knowledge as highly life-related through the use of this interactive mechanism.
**Kinesthetic clicking**

The KELS provided a virtual pointer that actively correlates the position of the learners’ hands with the position of the cursor on the computer screen. In order to trigger the system function by clicking the icon, learners moved the virtual pointer to the top of the icon via kinesthetic motion. In addition, a blue circle was designed to count down in order to activate the function.

**Activity module**

**Vocabulary learning**

There were four subjects in the vocabulary-learning activity: action, positioning, objects, and others. The vocabulary in the action subject was about various motion patterns. The objects vocabulary was about physical objects, and the vocabulary in the positioning subject was about prepositions used to indicate relative position. The others subject contained vocabulary that did not fit into the other subjects (e.g., happy, sad, confidence). In this study, four mechanisms that required interactive body movements were designed to trigger the display of information of vocabulary in four subjects, the corresponding physical image of the above four subjects in KELS are shown in Fig. 3. The motion mimic mechanism utilized the learners’ body movements to trigger the display of vocabulary in the “ACTION” subject. The display of information about the vocabulary in the “OBJECT” subject was triggered by the object recognition mechanism. Likewise, the gesture-based relative positioning mechanism triggered the display of information regarding prepositions in the “POSITION” subject. The kinesthetic clicking mechanism was used to display information about vocabulary in the “OTHER” subject. The detailed information and physical images of KELS displayed for learning vocabularies included spelling, a phonogram, and sample sentences (see the left side of Fig. 3). Advanced functions, including the pronunciation of vocabulary or example sentence using text to speech (TTS) and switching between multi-sample sentences (see the right side of Fig. 3), were also implemented.

![Figure 3. Menu of vocabulary learning (screenshot KELS)](image3)

![Figure 4. Interface of vocabulary information (screenshot KELS)](image4)
**Listening test**

In this stage, learners repeatedly listened to English questions provided by KELS and responded to the listening stimulation with respective body motions. Learners were able to understand and internalize the input through the interaction between listening stimulation and body motion. An English question was generated randomly by KELS and transmitted through TTS to learners, and they were required to respond to the question using the body interactive mechanism. The interface is shown in Fig. 5.

![Figure 5. Interface of listening test](image)

**Paired interactive speaking game**

In this stage, the game mode was paired cooperation. In each group, there were two learners, each using a separate computer. A socket connection was built between the two computers. KELS randomly generated a question to one of the learners in each pair, and the other learner was given four possible options from which to choose the correct answer. The learner with the question read it aloud through the microphone, and the other learner determined the answer from four options and made the selection using body motion. The two learners then exchanged roles for each subsequent question. The game was designed to train learners to speak English confidently as well as to hear and understand words as they were pronounced by their partners (Figure 6).

![Figure 6. Paired interactive speaking game](image)
Motion-based sentence making

Stevick (1976) investigated the relationship between memory and vocabulary development from the perspective of second language learning. He believed that active involvement was required in order to retain a vocabulary word in the brain for a long time. In other words, using the vocabulary is more effective to the learning process than just listening to, writing, or reading the translation of a given word. Thus, the learning activity at the last stage of the study required learners to make a sentence using body motion. KELS provided the learner with a picture, and the learner was required to respond by making a sentence describing the picture, causing the learner to retrieve vocabulary using the four body interactive mechanisms.

Method

Research participants and experimental time

This study was conducted with 39 tenth-grade students. The participants were selected from one class and were all female. They were divided into a control group and an experimental group, and the number of students in each group was 20 and 19, respectively. The teaching content for both groups was “Studio Classroom September.” Learning procedures were designed so that students read the teaching content during four sessions per week; each session was 30 minutes. The teaching lasted three weeks for a total of twelve sessions.

Research architecture

This study adopted a quasi-experimental method using the proposed KELS as the independent variable, perceptual style as the background variable, and learning achievements as dependent variables. The study further explored the relationship between system usage and learning achievements in the experimental group. The research architecture is shown in Fig. 7.
Activities design

This study was designed to investigate the differences in learning performance between the control group and the experimental group. Three major activities were prepared as follows:

The learning cycle for mastering learning

In this activity, the learning cycle was designed so that each learner went through the process of learning vocabulary and then took a listening test to see whether the vocabulary learning should continue or not. The learning cycle would continue until the participant passed the listening test, indicating the participant had mastered the subject matter. The vocabulary learning was to teach learners vocabulary words, sentences and pronunciation either through a multimedia site (control group) or through KELS (experimental group).

Paired interactive speaking activity

In this activity, one learner in a pair was required to speak while the other learner was required to respond. In the control group, learners were paired and two examination papers were distributed, paper A and paper B. Paper A contained a list of vocabulary words for one learner to read aloud, and paper B contained 4 vocabulary words, one of which corresponded to the vocabulary words listed in paper A, that is, the bearer of paper B was expected to select from the list the correct word based on what the bearer of paper A spoke. The roles were then exchanged. The activity was conducted similarly in the experimental group except that the vocabulary words were provided by KELS instead of by reading paper A and that the response was expressed by body interactive mechanisms instead of by reading paper B.

Sentence-making with pictures

In this activity, a picture was provided that the learner was required to describe in sentences using the vocabulary words that had been taught. The control group made sentences using paper and pencils while the experimental group used body motions to generate vocabulary to make sentences. The body motions were used to increase the learners’ memory and internalization in regard to learning contents.

Results

Analysis of learning effects

Before the experiment was conducted, we examined the learners’ prior knowledge of English in both the control group and the experimental group by using an independent-sample T-test. The results are shown in Table 1. There was no significant difference in prior knowledge between the groups. At the end of the three-week experiment, the post-test scores in the control group and the experiment group increased from the test of prior knowledge; however, there was still no significant difference in the scores. But, while the mean of pre-test scores in the control group was a lot higher than in the experiment group, the post-test scores in the control group was lower than in the experiment group. Also, the t-value changed from pre-test 1.678 to post-test -.589. Further, the difference of SD value between control group and experiment group shrank from pre-test to post-test. These were all clues indicating that the progress in the experiment group was better than in the control group. We also analyzed the learning gains and found that the experimental group was significantly better than the control group. A delay test was conducted 21 days after the learning activity, and the results showed that the experimental group performed significantly better than the control group on retaining what they had learned during the experiment.

Table 1. The learning effect by independent-sample t-test

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>20</td>
<td>52.60</td>
<td>12.339</td>
<td>1.678</td>
<td>.102</td>
</tr>
<tr>
<td>Experimental</td>
<td>19</td>
<td>44.89</td>
<td>16.172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
<td>-.589</td>
<td>.560</td>
</tr>
<tr>
<td>Control</td>
<td>20</td>
<td>67.80</td>
<td>11.998</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We intended to investigate the effect of system usage on learning effect and study the relationships among different system usages.

**Analysis of the relationship between system usage and post-test effects**

*Pearson’s Relation Analysis between Post-test Effect and the System Usage.* The average score and the highest score in paired interactive speaking activity were computed using scores from six learning sessions. These two variables had a positive correlation to the post-test; the correlation coefficients were .501 (p = .029 < .05) and .485 (p = .035 < .05) respectively. There are two possible interpretations or assumptions that can be drawn based on these results. First, if learners were equipped with better English abilities before the experiment, then after participating in all the learning sessions of the paired interactive speaking activity, they would have better learning effects in the post-test than learners with less English abilities. Second, learners who performed well in the paired interactive speaking activity through their practice in this stage might perform better in the post-test learning effects than learners performed worse in the paired interactive speaking activity. To clarify this question, we further examined the relation between the pretest score, the average score of the paired interactive speaking game, and the highest score of the paired interactive speaking game. We examined the relation of the pretest score to both the average score of the paired interactive speaking game and the highest score of the paired interactive speaking game through Pearson’s correlation analysis and found that there were no significant relations (p = .166 > .05; p = .184 > .05). As a result, the first assumption was excluded, and the second assumption was determined to be more likely.

*The Forecasting Ability of the System Usage for Post-test Effects.* Stepwise regression analysis was conducted to investigate the explanation of variance and to determine whether either the average or the highest score of the paired interactive speaking game corresponded to post-test effects. The result showed that the highest score of the paired interactive speaking game was eliminated. In addition, the explanation of variance of the joint regression model for the average score of the paired interactive speaking game (.207) revealed the explanatory power of 20.7%, F = 5.691 and p=.029, from which the difference was significant. Thus, the average score of the paired interactive speaking game could effectively predict the post-test score. In other words, learners who performed well in the paired interactive speaking game were likely to have better learning effects.

**Analysis of system usage on retaining effect obtained by the delay test**

*The Pearson’s Relation Analysis between System Usage and the Retaining Obtained by the Delay Test.* The total learning time refers to the total time learners spent in the vocabulary learning stage. The total test number refers to the total number of listening tests the learners took. The Pearson’s correlation analysis showed that both the retaining effect obtained by the delay test and the total learning time appeared in negative correlation with the correlation coefficient -.468 (p = .043 < .05) and that both the retaining effect obtained by the delay test and the total test number appeared in positive correlation with the correlation coefficient .463 (p = .046 < .05). There are two possible interpretations or assumptions that can be drawn based on these results. First, learners who spent less time on vocabulary learning and more time in taking the listening test might experience better learning effects; conversely, learners who spent more time on vocabulary learning and less time in taking the listening test might experience worse learning effects. Second, learners’ prior knowledge level might affect the results. The learning cycle for mastery was designed for vocabulary learning in which learners were examined by taking a listening test. Learners who did not perform well on the listening test would go through the vocabulary learning again. In this situation, learners with less prior knowledge of English spent more time on vocabulary learning than those whose prior

| Learning gains | Experimental | 19 | 70.11 | 12.463 |
| Control       | 20 | 15.20 | 11.547 | -.2517 | .016* |
| Learning retention | Experimental | 19 | 25.21 | 11.093 |
| Control       | 20 | 82.80 | 12.743 | -.2133 | .040* |

*p < .05. *p < .01. **p < .001.**
knowledge of English was greater. Alternately, learners with greater prior knowledge spent more time taking the listening tests.

To examine the assumptions mentioned above, we examined the relation between the pretest score and the delay test score. The result showed that the pretest and delay test scores had a greatly positive correlation; the correlation coefficient was $0.639 (p = .003 < .01)$. As a result, we concluded that the second assumption was more likely. The average score of the paired interactive speaking activity and the delay test scores appeared to have a positive correlation; the correlation coefficient was $0.527 (p = .020 < .05)$. Therefore, we concluded that the paired interactive speaking activity did help learners with long-term retention of their acquired knowledge.

*The Forecasting Ability of System Usage to the Retaining Effect Obtained by the Delay Test.* Stepwise regression analysis was conducted to investigate the explanation of variance on the retaining effect obtained by the delay test from the system usage which was total learning time, the total test number, and the average score of the paired interactive speaking activity. The result showed that the total learning time and total test number were not able to forecast the extent of long-term retention. The explanation of variance of the joint regression model for the average score of the paired interactive speaking activity was $0.236$, from which the explanatory power of $23.6\%, F = 6.547$ and $p = .020$ was significantly different. Thus, the average score of the paired interactive speaking activity effectively forecasted the extent of long-term retention. In other words, learners who performed well in the paired interactive speaking activity were equipped with better long-term retention of the English language.

**Questionnaire analysis**

*TAM technology acceptance model*

The Cronbach alpha value of the three dimensions in the technology acceptance model (TAM) questionnaire were all higher than 0.7, indicating that the questionnaire was reliable (perceived ease-of-use .848, perceived usefulness .902, user intention .727). The average score of perceived ease-of-use was 3.8 in which the score for object recognition function-related questions were particularly lower than for other functions. As a result, we were made aware that the learners did not like using the object recognition module because the recognition precision was not accurate or easy to use. The average score of perceived usefulness was 3.7. The scores of the object recognition function-related questions were in the range between 3.3 and 3.6, which were lower than other functions. However, upon interview, learners mentioned that:

> They were able to deepen the impression of the vocabulary by the action of taking up objects in object recognition and it was easier to learn while comparing to the Kinesthetic clicking.

Even though learners thought the object recognition function did help them acquire knowledge, improvement in both the precision and the speed of recognition was needed; therefore, scores were lower regarding this function. The average score of user intention was 3.9 in which the average score of each question was higher than 3.6, indicating that the learners had positive intention.

*Keller’s ARCS model of motivation design*

The ARCS questionnaire was given to the study participants to determine whether the KELS would benefit their motivation and willingness to learn. The Cronbach alpha value was higher than .7 for all four dimensions (attention: .918; relevance: .804; confidence: .802; and satisfaction: .847), which indicated that the responses to the questionnaire were reliable. The average attention score was 3.9. In interview, Learners mentioned that

> I like Kinect and prefer to learn by it because it is more interesting.
> I will think how to do it by using body action.

This showed that learners thought it was very interesting to operate by body action and it could effectively promote their attention and ensure their participants. The average relevance score was 3.7. The second item (“I have enough
time to learn in the KELS.”) and the third item (“I feel the learning contents in KELS is practical and related to my life.”) received lower scores than the other items on the questionnaire. Some of the learners thought that “I feel it waste time in listening test to share my time with partner to test” so they thought the learning time was not enough. In interview, Learners also mentioned that “I don’t have to use English in general days so I feel it have few relation with life.” and this might be the cause that learners thought the KELS had less relation with life. The average confidence score was 3.7. The second item (“I feel that KELS give back scores fairly.”) and the third item (“I am confident that I can learn the contents taught in the system well.”) received lower scores than the other items. Based on the responses given during the interview process and on the questionnaires, we discovered that learners thought their level of proficiency in English was not accurately reflected because a portion of their final scores relied upon how well their partners were able to speak the questions in the paired interactive speaking activity. Learners also reported that although they made progress in learning English using KELS, they still felt resistant to and less confident about learning EFL well because of their past experiences with learning using traditional teaching methods. Some learners found KELS more interesting than the traditional methods but still expressed a general dislike of learning EFL. The average satisfaction score was 3.8. In summary, the average score was higher than 3.6 for all items on the questionnaire, and the average score in each dimension was higher than 3.7, indicating that the learners did experience a greater motivation to learn EFL after using the KELS.

Analysis of perception learning style for learning

We intended to investigate whether the non-kinesthetic learners’ attention and retention of knowledge were negatively affected by engaging in highly physical activities that did not present knowledge using their primary perception learning style. Conversely, we also investigated whether the kinesthetic learners benefited from the use of body motion in the KELS. Although we discovered that kinesthetic learners performed slightly better than non-kinesthetic learners on both the learning test given immediately after the experiment and the delay test given 21 days after the experiment (see Table 2), there was no significant difference in the analysis. Based on the responses given during the interview process, non-kinesthetic learners demonstrated their appreciation with the learning mode presented in the KELS and expressed high motivation to continue their participation. In summary, whether learners were kinesthetic or non-kinesthetic, they were able to use and learn through the KELS well.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-K</td>
<td>11</td>
<td>69.45</td>
<td>13.178</td>
<td>-.260</td>
<td>.798</td>
</tr>
<tr>
<td>K</td>
<td>8</td>
<td>71.00</td>
<td>12.236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning gains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-K</td>
<td>11</td>
<td>89.73</td>
<td>14.907</td>
<td>-.696</td>
<td>.496</td>
</tr>
<tr>
<td>K</td>
<td>8</td>
<td>93.88</td>
<td>9.078</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05. ** p < .01. *** p < .001.

The discussion and educational implication

In summary, the stepwise regression analysis indicated that the paired interactive speaking activity was better able to forecast the results on the post-test and predict the extent the learning would be retained as indicated on the delay test (for long-term retention). This ability coincides with the theorems stating that learning is more effective when interaction with other people is incorporated, as opposed to using videos or tapes, because the learning occurred in a cultural, practical, and meaningful context. Using partner interaction supported the effort to engage the learners’ brains by presenting the knowledge as practical and life-related; therefore, the participants learned more quickly and retained the knowledge for a longer than they were likely to have in a traditional classroom setting.

Open questions and interviews were analyzed to determine any correlations between the responses and our findings: The paired interactive speaking activity was found to be important and beneficial to learning EFL. The interviews indicated that most learners preferred the paired interactive speaking activity not only because it was more interesting with two members of a group cooperating together but also because hearing the sentences spoken by the
partner resulted in better learning than listening to the TTS system alone. In this activity, the learners tended to speak English more confidently.

The results of the ARCS questionnaire indicated that the learners did experience a greater motivation to learn EFL after using the KELS. The results of learning style analysis showed that whether learners were kinesthetic or non-kinesthetic, they were able to use and learn through the KELS well.

Conclusions

In Taiwan, the focus on reading and writing skills in the EFL learning environment neglects the importance of the learning order: first mastering listening comprehension, then speaking, and lastly reading and writing; this approach also requires more effort from learners while reducing the effectiveness of the learning process. Because the learning mostly takes place in the classroom rather than within life-related situations, learners experience less confidence and interest, resulting in lower motivation and learning retention. In contrast to this situation, many studies indicated that interaction with the subject matter using bodily motion could enhance problem-solving and make learning more effective. In addition, using bodily motion to indicate comprehension on a listening test may improve the brain’s ability to link the knowledge acquired with life-related tasks and thus increase the effectiveness of the learning. Therefore, this study incorporated TPR in the proposed Kinesthetic English Learning System (KELS) using Microsoft Kinect technology. The system consisting of four stages of learning activities and four body interactive mechanisms, was designed for high school seniors as a learning tool.

There were no significant differences between the control group and the experimental group on the learning test taken at the conclusion of the experiment; however, the experimental group performed better on the delay test that was given 21 days later to measure long-term retention. We concluded that the KELS effectively enhanced the internalization of EFL learning and promoted long-term retention, increasing the likelihood that learners will be able to apply what they have learned to daily life. In the four stages of the KELS, learners examined themselves and repeatedly continued the body motion to response to Vocabulary Learning and Listening Test. The interactive mode helped learners to better internalize knowledge. We also found that the paired interactive speaking activity promoted long-term retention of EFL. Through the TAM questionnaire, ARCS questionnaire, open questionnaire and interview process, we discovered that most learners in the experimental group thought that the kinesthetic learning mode incorporating body motion was interesting and held their attention. In particular, using body motion to express the meaning of words made the learning process more effective and increased the association of EFL with life-related tasks. Although several learners expressed initial resistance to learning English because of their previous negative experiences with the subject, they did acknowledge that their knowledge increased throughout the experiment while the sense of resistance decreased progressively. Overall, the study participants were equipped with high motivation to continue participating in the KELS. Finally, the results of the experimental group indicated that there was no significant difference between kinesthetic and non-kinesthetic learners in either learning or retaining knowledge.

The limitation of this work is that the number of students participated the experiment is limited and they are with similar cultural background. As a result, in the future work, a large scale of subject test with more sample sizes is on the way in order to further verify KELS. Besides, context factors, such as gender, age, profession or cultural background, will be considered.

References


